

Best Local Similarity 100.0%; Pred. No. 1.83e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 atqklytrisygid 15
| | | | | | | | | |
OY 1 ATOKITRISGVGID 15

RESULT 2
ID M07908 standard; protein: 614 AA.
AC M07908;
DE 29-JAN-1997 (first entry)
KW Pemphigus vulgaris antigen protein extracellular region.
KW Autoantibody; Immunoglobulin G; IgG1; fusion protein; diagnosis;
KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;
KW dermatology.
OS Homo sapiens.
PN J08188540-A.
PD 23-JUL-1996.
PF 30-JUN-1995; 165632.
PR 30-JUN-1994; JP-173291.
PA (NISH/) NISHUKAWA T.
DR WPI: 96-388562/39.
PT Fused protein recognised by pemphigus vulgaris autoantibody -
PT useful to treat and diagnose pemphigus vulgaris
PS Claim 1; Page 7-9; 9pp; Japanese.
CC M07908 represents the human pemphigus vulgaris (PV) antigen
CC extracellular region. The PV antigen is produced in patients with
CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare
CC relapsing disease causing suprabasal, intra-epidermal bullae
CC (vesicles) of the skin and mucous membranes, which is fatal if
CC untreated. The PV antigen was fused to a human IgG1 hinge region
CC and the resulting fusion protein is useful to treat or diagnose
CC pemphigus vulgaris.
SQ Sequence 614 AA;

Query Match 100.0%; Score 101; DB 19; Length 614;
Best Local Similarity 100.0%; Pred. No. 1.83e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 78 atqklytrisygid 92
| | | | | | | | | |
OY 1 ATOKITRISGVGID 15

RESULT 3
ID R30742 standard; protein: 999 AA.
AC R30742;
DE 14-JUN-1993 (first entry)
KW Human pemphigus vulgaris 130kD antigen.
KW Pemphigus vulgaris; skin disease; autoantibodies;
KW keratinocyte cell surface antigen; glycoprotein; cell adhesion.
OS Homo sapiens.
PN US7798918-A.
PD 15-DEC-1992.
PF 27-NOV-1991; 798918.
PR 27-NOV-1991; US-798918.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Amagai M, Klaus-Kovtun V, Stanley JR;
DR WPI: 93-067436/08.
DR N-PSDB: 035992.
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
PT diagnostic and therapeutic uses
PS Disclosure; Fig 7; 50pp; English.
CC This sequence is the pemphigus vulgaris 130kD antigen. The protein
CC and its encoding DNA may be used in the diagnosis and treatment of
CC pemphigus vulgaris. It is thought that the antigen may be a cell
CC adhesion molecule.
SQ Sequence 999 AA;

Query Match 100.0%; Score 101; DB 6; Length 999;
Best Local Similarity 100.0%; Pred. No. 1.83e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 79 atqklytrisygid 93
| | | | | | | | | |
OY 1 ATOKITRISGVGID 15

RESULT 4
ID W15489 standard; protein: 778 AA.
AC W15489;
DE 17-JUN-1997 (first entry)
KW Pemphigus foliaceus antigen-IgG constant region fusion protein.
KW Pemphigus foliaceus; autoantibody; constant region; IgG;
KW extracellular region; antigen; hinge portion; skin;
KW dermatitis herpetiformis; fusion protein; detection; ss.
OS Chimeric - Homo sapiens.
FH Key
FT domain Location/Qualifiers
FT 1..545 /note="Pemphigus foliaceus antigen protein"
FN J09077800-A.
PD 25-MAR-1997.
PE 12-SEP-1995; 260899.
PR 12-SEP-1995; JP-260899.
PA (NISH/) NISHUKAWA T.
DR WPI: 97-241756/22.
DR P-PSDB: T66428.
PT Pemphigus foliaceus antigen-IgG constant region fusion protein - linked
PT through the hinge region used to treat pemphigus foliaceus
PS Claim 1; Page 10-12; 17pp; Japanese.
CC This sequence represents a fused protein recognised by pemphigus
CC foliaceus patient autoantibody which comprises the constant region
CC of IgG linked to the extracellular region of pemphigus foliaceus
CC antigen protein through the hinge portion. Pemphigus foliaceus is
CC a chronic, generalised, vesicular and scaling skin eruption similar
CC to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion
CC protein is useful to treat pemphigus foliaceus. The antigen is
CC especially administered through an adsorbent upon which the fusion
CC protein is immobilised via a carrier. The fusion protein is also
CC useful for detecting pemphigus foliaceus antibodies which is also
CC in immunodiagnosis. The fusion protein has little or no side effects.
SQ Sequence 778 AA;

Query Match 86.1%; Score 87; DB 21; Length 778;
Best Local Similarity 80.0%; Pred. No. 9.69e-03;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 79 angvtrisygid 93
| | | | | | | | | |
OY 1 ATOKITRISGVGID 15

RESULT 5
ID W31706 standard; protein: 1363 AA.
AC W31706;
DE 14-APR-1998 (first entry)
KW Bovine coronavirus E2 (S) protein.
KW BCV; E2 protein; peplomer protein; S spike; antigen; vaccine;
KW cattle.
OS Bovine coronavirus.
FH Key
FT Peptide Location/Qualifiers
FT 1..17 /label="S1g_peptide"
FT Protein 18..1363 /label="Mat_protein"
FT 1306..1338 /label="Mat_protein"
FT Domain /note="transmembrane domain"
FN US5672350-A.
PD 30-SEP-1997.
PF 22-AUG-1989; 397689.
PR 19-DEC-1991; US-811422.
PR 22-AUG-1989; US-397689.
PR 18-OCT-1991; US-779500.
PR 22-DEC-1993; US-171763.
PA (VETE-) VETERINARY INFECTIOUS DISEASE.

PI Babluk LA, Cox GJ, Parker MD;
 DR WPI: 97-488823/45.
 DR N-PSDB: T89387.
 PT Vaccines against bovine coronavirus - containing recombinant bovine
 coronavirus polypeptide(s)
 PS Claim 13: Fig 3: 53pp; English.
 CC This polypeptide comprises the E2 protein, also designated
 CC peplomer protein or S (Spike), of bovine coronavirus (BCV). It
 CC has a mol.wt. of 150 kDa exclusive of glycosylation and contains
 CC 21 potential N-linked glycosylation sites. The amino acid
 CC sequence was deduced from an clone E2 cDNA (see T89387). The
 CC E2 gene in plasmid pT18E2 (E. coli JM105) is deposited as ATCC
 CC 68041. The BCV E3 gene (see T89388) is immediately 5' of the E2
 CC gene on the viral genome and terminates 14 nucleotides upstream
 CC from the E2 initiation codon. The E2 and E3 genes have been
 CC cloned and can be used for the recombinant production of BCV
 CC polypeptides, using e.g. Spodoptera frugiperda Sf9 insect cells
 CC as host cells. Glycosylated and non-glycosylated recombinant
 CC E2 and E3 (see W31707) are useful as components of vaccines
 CC directed toward preventing BCV infection, or reducing the severity
 CC of BCV infection, in bovine populations.
 CC Sequence 1363 AA;

Query Match 56.4%; Score 57; DB 27; Length 1363;
 Best Local Similarity 45.5%; Pred. No. 2.94e+01;
 Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 995 nvqyringiv 1005
 OY 4 KITRYISGVGI 14

RESULT 6
 ID R1061 standard; Protein; 1363 AA.
 AC R11061;
 DT 24-MAY-1991 (first entry)
 DE Bovine Coronavirus E2 protein.
 KW Bovine Coronavirus; BCV; E2; E3; vaccine.
 GS Bovine Coronavirus - Quebec isolate.
 FH Key Location/Qualifiers
 FT Peptide 1..17
 FT /label= sig_peptide
 FT cleavage_site 766..766
 FT /note= "probable cleavage site"
 FT domain 1306..1338
 FT /label= transmembrane_domain
 PN MO9102752-A.
 PD 07-MAR-1991.
 PF 10-AUG-1990; CA0252.
 PR 22-AUG-1989; US-397689.
 PA (VETE-) VETERINARY INFECTION.
 PI Parker MD, Cox GJ, Babluk LA;
 DR WPI: 91-087247/12.
 DR N-PSDB: 010947.
 PT Decyribonucleic acid encoding Bovine coronavirus protein E2 and
 PT E3 - useful as vaccine component
 PS Disclosure; Fig 3 (1-6); 70pp; English.
 CC E2 protein is one of the four proteins (N, E1, E2 and E3) composing
 CC BCV. The E2 and E3 polypeptides, or fragments of these, are used as
 CC subunit antigens in vaccines for protection against or ameliorating
 CC BCV, without risk of infection.
 CC Sequence 1363 AA;

Query Match 56.4%; Score 57; DB 2; Length 1363;
 Best Local Similarity 45.5%; Pred. No. 2.94e+01;
 Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 995 nvqyringiv 1005
 OY 4 KITRYISGVGI 14

RESULT 7

ID R85487 standard; Protein; 878 AA.
 AC R85487;
 DT 18-MAR-1996 (first entry)
 DE Human E-cadherin precursor.
 KW E-cadherin; T-lymphocyte; alpha-E,beta-7 integrin; cell adhesion;
 KW autoimmune disease; Crohn disease; psoriasis.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..150
 FT /label= sig_peptide
 FT domain 151..702
 FT /label= Extracellular domain
 FT /note= "the extracellular domain (amino acids 1-552
 FT of the mature protein) is the preferred
 FT region for generation of peptides of the
 FT invention"
 FT domain 703..726
 FT /label= Transmembrane_domain
 FT domain 727..876
 FT /label= Cytoplasmic_domain
 PN MO9529693-A1.
 PD 09-NOV-1995.
 PF 03-MAY-1995; U05518.
 PR 03-MAY-1994; US-237919.
 PA (BGHM) BRIGHAM & WOMENS HOSPITAL.
 PI Brenner MB, Cepex KL;
 DR WPI: 95-392921/50.
 DR N-PSDB: T05764.
 PT Inhibiting adhesion of T lymphocytes with E-cadherin - useful for
 PT isolating agents to treat auto-immune diseases e.g. Crohn's disease,
 PT psoriasis, etc
 PS Disclosure; Page 70-75; 103pp; English.
 CC The human E-cadherin protein precursor (R85487) is expressed by
 CC an cDNA clone (T05764) derived from human liver. The extracellular
 CC domain of E-cadherin is used to generate peptides that specifically
 CC bind to heterotypic cognates of E-cadherin and which inhibit adhesion
 CC of intra-epithelial T-lymphocytes to E-cadherin-expressing epithelial
 CC or endothelial cells in vitro or in vivo, thereby modulating
 CC mucosal immune responses. Such peptides are also specifically
 CC reactive with a monoclonal antibody (E4.6 or E6.1) that binds to
 CC E-cadherin and that can inhibit T-cell binding.
 CC Sequence 878 AA;

Query Match 53.5%; Score 54; DB 15; Length 878;
 Best Local Similarity 100.0%; Pred. No. 6.20e+01;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 515 qkityri 521
 OY 3 QKITRYRI 9

RESULT 8
 ID R55060 standard; Protein; 878 AA.
 AC R55060;
 DT 08-NOV-1994 (first entry)
 DE Sequence of human liver E-cadherin.
 KW Human epithelial-cadherin; E-cadherin; cell adhesion molecule; CAM;
 KW uvomoulin; L-CAM; Cell CAM 120/80.
 OS Homo sapiens.
 PN MO9411401-A.
 PD 26-MAY-1994.
 PF 16-NOV-1993; U11097.
 PR 17-NOV-1992; US-978897.
 PA (UYVA) UNIV YALE.
 PI Morrow JS, Rimm DL;
 DR WPI: 94-183426/22.
 DR N-PSDB: 065487.
 PT Purified human E-cadherin protein and nucleic acid - used to
 PT develop prods. for diagnosis, prognosis, therapy and prophylaxis
 PT of t-cadherin disorders, e.g. malignancies
 PS Claim 1; Page 59-63; 97pp; English.
 CC E-cadherin is a cell adhesion molecule that is also known as

CC uvomomulin, L-CAM and Cell CAM 120/80. The DNA encoding hbc was obtd.
 CC by screening normal human liver and hepatocellular carcinoma cDNA
 CC libraries and a colonic epithelial cell cDNA library. The following
 CC sequences are specifically claimed: Aas 1-878; 151-878; 30
 CC sequential Aas from Aas 308-878; Aas 1-150; Aas 178-289; Aas 290-
 CC 401; Aas 402-513; Aas 178-513; Aas 151-703; Aas 1-703; Aas 728-878;
 CC Aas 704-878; nucleotide sequences comprising nucleotide numbers 116-
 CC 2749; 566-2749; 1-1053; 510-2696; 1332-3000; 540-1500; 348-906; 890-
 CC 1548; 384-1208; 641-2046; 685-1336; 880-1661; 1199-1742; 1373-1742;
 CC 1705-2204; 2458-2775; DNA encoding at least 30 Aas selected from
 CC Aas 308-878. The prods. can be used in the diagnosis, prognosis,
 CC therapy and prophylaxis of conditions involving improper E-cadherin
 CC expression. Suitable dosages for i.v. admin. of a protein are
 CC 20-500 mcg/kg body wt.
 CC Sequence 878 AA;

Query Match 53.5%; Score 54; DB 10; Length 878;
 Best Local Similarity 100.0%; Pred. No. 6.20e+01;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 515 qklyr1 521
 |||||
 QY 3 QKTYR1 9

RESULT 9
 ID R06441 standard; protein: 309 AA.
 AC R06441;
 DT 08-OCT-1990 (first entry)
 DE MG-6 antigen.
 KW Mycoplasma gallisepticum; poultry; vaccine.
 PN J02111795-A.
 PD 24-APR-1990.
 PE 02-JUN-1989; 136343.
 PF 02-JUN-1989; JP-136343.
 PA (JAFG) Nippon Zeon KK, (SHIO) Shionogi KK.
 DR WPI; 90-169109/22.
 DR N-PSDB; 005653.
 PT Diagnostic and vaccine for poultry mycoplasma serum - utilises
 PT antigen protein of the disease and recombinant vector
 FI Incorporated with its coding gene.
 CS Claim 2; Fig 1f; 20pp; Japanese.
 CC DNA encoding the protein can be inserted into an expression vector
 CC for the prodn. of MG-1 polypeptide which elicits an antigen-antibody
 CC reaction with anti-mycoplasma gallisepticum poultry sera. It may also
 CC be ligated to other DNA to produce fusion proteins with an N-terminal
 CC bacterial enzyme sequence.
 CC See also R05081-2 and R06437-40.
 CC Sequence 309 AA;
 SQ

Query Match 52.5%; Score 53; DB 1; Length 309;
 Best Local Similarity 58.3%; Pred. No. 7.93e+01;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 200 tqvtvntfagvg 211
 ||:|:|:|
 QY 2 TOKITRYISGVG 13

RESULT 10
 ID W37140 standard; Protein: 626 AA.
 AC W37140;
 DT 06-JUL-1998 (first entry)
 DE Pea plastidial phosphoglucosyltransferase.
 KW Phosphoglucosyltransferase; pea; PGM(p); sucrose; transgenic plant;
 KW ring3 mutation.
 OS Pisum sativum var. Novella.
 PN W09801574-A1.
 PD 15-JAN-1998.
 PE 03-JUL-1997; E03613.
 PF 10-FEB-1997; GB-002653.
 PR 09-JUL-1996; US-021410.
 PR 18-JUL-1996; GB-015103.

PA (UNITL) UNILEVER NV.
 PA (UNITL) UNILEVER PLC.
 PI Harrison CJ, Hedley CL, Hugues SG, Lacey CND, Wang TL;
 DR WPI; 98-101065/09.
 DR N-PSDB; V00731.

PT Pea seeds lacking plastidial phosphoglucosyltransferase activity - have
 PT higher sucrose levels at end of vining period and can be vined over
 PT extended period compared to conventional varieties
 PS Disclosure; Page 58-62; 83pp; English.
 CC This protein comprises pea plastidial phosphoglucosyltransferase (PGM(p)).
 CC an enzyme that catalyses the interconversion of glucose-1-phosphate
 CC and glucose-6-phosphate, and is thus involved in starch synthesis.
 CC The amino acid sequence was deduced from an isolated cDNA clone
 CC (see V00731) and shows considerable homology to known PGM sequences.
 CC The invention provides pea plants that have substantially reduced
 CC PGM(p) activity or which substantially lack PGM(p) activity. This
 CC may be the result of antisense or sense suppression technology.
 CC Pea seeds from such plants have higher sucrose levels at the end of
 CC the vining period and may be vined over an extended period compared
 CC to conventional varieties. The invention provides pea seeds having
 CC a sucrose content of over 6% of total weight, and a ratio of
 CC sucrose content to start content of over 10.
 CC Sequence 626 AA;
 SQ

Query Match 51.5%; Score 52; DB 30; Length 626;
 Best Local Similarity 60.0%; Pred. No. 1.01e+02;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 563 rilyrlsqtg 572
 :|:|:|
 QY 4 KITRYISGVG 13

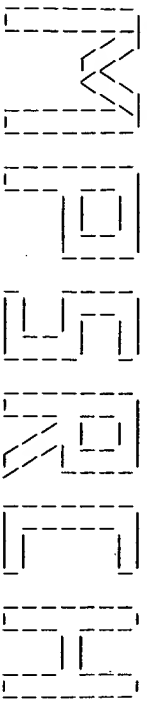
RESULT 11
 ID W73500 standard; protein: 504 AA.
 AC W73500;
 DT 26-FEB-1999 (first entry)
 DE Trabecular meshwork induced glucocorticoid response protein.
 KW TIGR protein; trabecular meshwork induced glucocorticoid response;
 KW secretory protein; antibody; glaucoma; diagnosis.
 OS Homo sapiens.
 PN US5849879-A.
 PD 15-DEC-1998.
 PF 14-MAY-1996; 645900.
 PR 14-MAY-1996; US-645900.
 PR 03-NOV-1994; US-336235.
 PR 20-OCT-1995; US-546568.
 PA (REGC) UNIV CALIFORNIA.
 PI Huang W, Nguyen TD, Polansky JR;
 DR WPI; 99-069807/06.
 PT Antibody to trabecular meshwork protein - useful for diagnosis of
 PT glaucoma

PS Claim 1; Column 25-28; 22pp; English.
 CC This sequence represents the human trabecular meshwork induced
 CC glucocorticoid response (TIGR) protein. The TIGR protein is a secretory
 CC protein specifically bound by the antibody of the invention. The
 CC antibody, especially in labelled form, can be used in the diagnosis of
 CC glaucoma by detecting elevated levels of the protein in the trabecular
 CC meshwork of the eye. Using the antibody, glaucoma is detected more
 CC accurately.
 CC Sequence 504 AA;
 SQ

Query Match 50.5%; Score 51; DB 38; Length 504;
 Best Local Similarity 57.1%; Pred. No. 1.29e+02;
 Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 281 tqetwridtvgtd 294
 |||||
 QY 2 TOKITRYISGVGID 15

RESULT 12
 ID W89391 standard; Protein: 504 AA.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Jun 11 17:12:13 1999; MasPar time 4.23 Seconds
Tabular output not generated. 142.091 Million cell updates/sec

Title: >US-08-991-628-1
Description: (1-15) from US08991628.pep
Perfect score: 101
Sequence: 1 ATOKITRISGVID 15

Scoring table:
PAM 150
Gap 15

Searched: 122810 segs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: plr60
1:plr1 2:plr2 3:plr3 4:plr4

Statistics: Mean 25.979; Variance 34.700; scale 0.749

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	101	100.0	999	1	IJHUG3	desmoglein 3 precursor	4.47e-10
2	87	86.1	1043	1	IJBGG1	desmoglein 1 precursor	1.07e-06
3	87	86.1	1049	1	IJHUG1	desmoglein 1 precursor	1.07e-06
4	58	57.4	331	2	E70367	phosphoribosylformyl	1.99e+00
5	57	56.4	441	2	T01000	hypothetical protein	3.07e+00
6	57	56.4	1363	1	VG1HNM	E2 glycoprotein precu	3.07e+00
7	57	56.4	1363	1	VG1HNM	surface protein - hum	3.07e+00
8	57	56.4	1363	1	VG1HNM	E2 glycoprotein precu	3.07e+00
9	57	56.4	1363	1	VG1HNM	E2 glycoprotein precu	3.07e+00
10	57	56.4	1363	1	VG1HNM	E2 glycoprotein precu	3.07e+00
11	57	56.4	1363	1	VG1HNM	E2 glycoprotein precu	3.07e+00
12	57	56.4	1363	1	VG1HNM	surface protein - hum	3.07e+00
13	57	56.4	1363	1	VG1HNM	E2 glycoprotein precu	3.07e+00
14	55	54.5	108	2	J01428	hypothetical protein	4.74e+00
15	55	54.5	169	2	S73495	hypothetical protein	7.28e+00
16	55	54.5	169	2	S73495	hypothetical protein	7.28e+00
17	55	54.5	169	2	S73495	hypothetical protein	7.28e+00
18	55	54.5	169	2	S73495	hypothetical protein	7.28e+00
19	54	53.5	479	2	A39405	beta-galactosidase (E	1.11e+01
20	54	53.5	509	2	A35540	cruciferin 1 precursor	1.11e+01
21	54	53.5	531	2	S17937	xpse protein - Xantho	1.11e+01
22	54	53.5	882	1	IJHUCF	cadherin 1 precursor	1.11e+01
23	54	53.5	884	1	IJHUCF	E-cadherin precursor	1.11e+01

24	54	53.5	884	2	S34438	uvomorulin - mouse	1.11e+01
25	53	52.5	82	2	I51134	MHC class II beta cha	1.69e+01
26	53	52.5	82	2	I51135	MHC class II beta cha	1.69e+01
27	53	52.5	82	2	I51136	MHC class II beta cha	1.69e+01
28	53	52.5	280	2	S50369	probable membrane pro	1.69e+01
29	53	52.5	813	2	B47485	ABR protein 2 - human	1.69e+01
30	53	52.5	822	2	A47485	ABR protein 1 - human	1.69e+01
31	53	52.5	859	2	A49307	98k GRPase-activating	1.69e+01
32	53	52.5	890	2	A30481	bacteriocin BCNS - Cl	1.69e+01
33	53	52.5	940	2	H71409	hypothetical protein	1.69e+01
34	53	52.5	1177	2	D71416	probable PDR5-like AB	1.69e+01
35	53	52.5	1335	1	VG1HMT	E2 glycoprotein precu	1.69e+01
36	53	52.5	1353	1	J02168	E2 glycoprotein precu	1.69e+01
37	53	52.5	1361	2	S29998	surface protein - hum	1.69e+01
38	53	52.5	1362	2	A37474	surface glycoprotein	1.69e+01
39	53	52.5	1376	1	VG1HJ2	E2 glycoprotein precu	1.69e+01
40	53	52.5	1376	1	J01534	E2 glycoprotein precu	1.69e+01
41	52	51.5	177	2	S65159	hypothetical protein	2.55e+01
42	52	51.5	435	2	S68865	fibrillar protein hife	2.55e+01
43	52	51.5	575	2	JC4554	ABC-type transporter	2.55e+01
44	52	51.5	608	2	H71379	probable tpr protein	2.55e+01
45	52	51.5	1324	1	VG1H59	E2 glycoprotein precu	2.55e+01

ALIGNMENTS

RESULT 1
ENTRY IJHUG3 #type complete
TITLE desmoglein 3 precursor - human
ALTERNATE_NAMES pemphigus vulgaris antigen
ORGANISM #formal_name Homo sapiens #common_name man
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 18-Sep-1998

ACCESSIONS A41088
REFERENCE A41088
#authors Amagai, M.; Klaus-Kovtun, V.; Stanley, J.R.
#journal Cell (1991) 67:869-877
#title Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a disease of cell adhesion.

#cross-references MUID:92069753
#accession A41088

GENETICS
#gene GDB:DSG3
#map_position 18q12.1-18q12.2
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein

FEATURE 1-23
24-49
30-999
#domain signal sequence #status predicted #label SIG
#domain propeptide #status predicted #label PRO
#product desmoglein homolog #status predicted #label MAT
#domain extracellular #status predicted #label EXT
#domain cadherin repeat homology #label CR1
#domain cadherin repeat homology #label CR2
#domain cadherin repeat homology #label CR3
#domain cadherin repeat homology #label CR4
#domain transmembrane #status predicted #label TM
#domain intracellular #status predicted #label INT
#domain desmoglein repeat #label DG1
#domain desmoglein repeat #label DG2
#binding_site carbohydrate (Asn) (covalent) #status predicted

SUMMARY #length 999 #molecular-weight 107502 #checksum 8311

Query Match 100.0%; Score 101; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 4.47e-10;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 79 ATOKITRISGVGID 93
 |||||
 1 ATOKITRISGVGID 15

RESULT 2
 ENTRY 1B0G1 #type complete
 TITLE desmoglein 1 precursor - bovine
 ALTERNATE_NAMES desmoglein BDGM
 ORGANISM #normal_name Bos primigenius taurus #common_name cattle
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 05-Sep-1997

ACCESSIONS
 #authors S14603; A38872; A37785; S38721; A48173; S24412
 #submission Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
 #description Submitted to the EMBL Data Library, March 1991
 Complete sequence of the desmoglein precursor and evidence for the existence of different desmoglein genes expressed in cell type-specific patterns.

#accession S14603
 #molecule_type mRNA
 #residues 1-1043 #label KOC

REFERENCE
 #cross-references EMBL:X58466; NID:9306; PID:9307
 #authors Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.

#journal Eur. J. Cell Biol. (1991) 55:200-208
 #title Complete amino acid sequence of the epidermal desmoglein precursor polypeptide and identification of a second type of desmoglein gene.

#cross-references M01D:92037656
 #accession A38872
 #molecule_type mRNA
 #residues 1-87;968-1043 #label KO2

REFERENCE
 #cross-references GB:S64268; GB:S64270
 #authors A37785
 #journal Goodwin, L.; Hill, J.E.; Raynor, K.; Raszi, L.; Manabe, M.; Cowin, P.

#journal Biochem. Biophys. Res. Commun. (1990) 173:1224-1230
 #title Desmoglein shows extensive homology to the cadherin family of cell adhesion molecules.

#cross-references M01D:91097553
 #accession A37785
 #molecule_type mRNA
 #residues 44-123; 'V', 125-493 #label GOO

REFERENCE
 #cross-references GB:M58165; NID:9162966; PID:9552318
 #authors S38721
 #journal Zimbelmann, R.

#authors Submitted to the EMBL Data Library, February 1991
 #accession S38721
 #molecule_type mRNA
 #residues 44-1043 #label ZIM

REFERENCE
 #cross-references EMBL:X57784; NID:9436061; PID:9436062
 #authors A48173
 #journal Koch, P.J.; Walsh, M.J.; Schmelz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.

#journal Eur. J. Cell Biol. (1990) 53:1-12
 #title Identification of desmoglein, a constitutive desmosomal glycoprotein, as a member of the cadherin family of cell adhesion molecules.

#cross-references M01D:91168965
 #accession A48173
 #molecule_type mRNA
 #residues 44-1001 'AOPPSAT' #label KO3

GENETICS
 #gene DSG1
 #classification #superfamily cadherin; cadherin repeat homology
 #keywords calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein

FEATURE
 1-23
 24-49
 50-1043
 50-548
 52-157
 160-269
 272-385
 392-491
 548-574
 575-1043
 846-875
 876-905
 906-933
 934-962
 963-1012
 110
 180,496

SUMMARY
 #length 1043 #molecular-weight 112242 #checksum 6897

Query Match 86.1%; Score 87; DB 1; Length 1043;
 Best Local Similarity 80.0%; Pred. No. 1.07e-06;
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DB 79 ANOQVTRISGVGID 93
 |||||
 1 ANOQVTRISGVGID 15

RESULT 3
 ENTRY 1YH0G1 #type complete
 TITLE desmoglein 1 precursor - human
 ALTERNATE_NAMES desmosomal glycoprotein I
 ORGANISM #normal_name Homo sapiens #common_name man
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 26-Feb-1998

ACCESSIONS
 #authors S16906; A39706; A61254; A61279; S16158
 #submission Buxton, R.S.
 #accession S16906
 #molecule_type mRNA
 #residues 1-1049 #label BUX

REFERENCE
 #cross-references EMBL:X56654; NID:930505; PID:930506
 #authors A39706
 #journal Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Ataliotis, P.; Poynter, D.; Arnemann, J.; Rutman, A.J.; Pidsley, S.C.; Watt, F.M.; Rees, D.A.; Buxton, R.S.; Magee, A.I.

#journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:4796-4800
 #title Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions, is related to the cadherin family of cell adhesion molecules.

#cross-references M01D:91271279
 #accession A39706
 #molecule_type mRNA
 #residues 24-1049 #label WHE

REFERENCE
 #cross-references GB:X56654
 #authors A61254
 #journal Niles, L.A.; Parry, D.A.D.; Powers, E.E.; Angst, B.D.; Wagner, R.M.; Green, K.J.

#journal J. Cell Sci. (1991) 99:809-821
 #title Structural analysis and expression of human desmoglein: a cadherin-like component of the desmosome.

#accession A61254
 #molecule_type mRNA
 #residues 26-1049 #label NIL

REFERENCE
 #authors A61279
 #journal Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arnemann, J.; Rees, D.A.; King, I.A.; Magee, A.I.

#journal Biochem. Soc. Trans. (1991) 19:1060-1064
 #title Desmosomal glycoproteins I, II and III: novel members of the

cathearin superfamily.
#cross-references MUID:92175187
#accession A61279
#status not compared with conceptual translation
#molecule_type mRNA
#residues 1-55 ##label WH3

GENETICS
#gene GDB:DSG1
#cross-references GDB:126563; OMIM:125670
#map_position 18q12.1-18q12.2
CLASSIFICATION #superfamily cathearin; cathearin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;
transmembrane protein

FEATURE
1-23 #domain signal sequence #status predicted #label SIG/
24-49 #domain propeptide #status predicted #label PRO/
50-1049 #product desmoglein #status predicted #label MAT/
50-548 #domain extracellular #status predicted #label EXT/
52-157 #domain cathearin repeat homology #label CR1/
160-269 #domain cathearin repeat homology #label CR2/
272-385 #domain cathearin repeat homology #label CR3/
392-493 #domain cathearin repeat homology #label CR4/
509-550 #region serine/threonine-rich
549-569 #domain transmembrane #status predicted #label TMA/
572-1049 #domain intracellular #status predicted #label INT/
840-869 #domain desmoglein repeat #label DG1/
870-889 #domain desmoglein repeat #label DG2/
900-927 #domain desmoglein repeat #label DG3/
928-956 #domain desmoglein repeat #label DG4/
966-1019 #region glycine/serine-rich
110,180 #binding_site carbohydrate (Asn) (covalent) #status predicted

SUMMARY #length 1049 #molecular-weight 113715 #checksum 4482

Query Match 86.1%; Score 87; DB 1; Length 1049;
Best Local Similarity 80.0%; Pred. No. 1.07e-06;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DB 79 ANQOVTYRISGVGID 93
QY 1 ATOKITRISGVGID 15

RESULT 4
ENTRY E70367
TITLE phosphoribosylformylglycinamide cyclo-lyase - Aquifex
ORGANISM aeolicus
#formal_name Aquifex aeolicus
#formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change
16-Dec-1998
ACCESSION E70367
REFERENCE A70300
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;
Keller, M.; AuJay, M.; Huber, R.; Feldman, R.A.; Short,
J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium
Aquifex aeolicus.
#cross-references MUID:98196666
#accession E70367
#status preliminary; nucleic acid sequence not shown;
translation not shown

GENETICS
#molecule_type DNA
#residues 1-331 ##label AOF
#cross-references GB:AE000706; NID:92983327; PID:92983329; GB:AE000657
#experimental_source strain VFS

CLASSIFICATION #superfamily phosphoribosylformylglycinamide cyclo-lyase;
phosphoribosylformylglycinamide cyclo-lyase homology
FEATURE
4-317 #domain phosphoribosylformylglycinamide cyclo-lyase

homology #label PfCL
SUMMARY #length 331 #molecular-weight 36768 #checksum 7544

Query Match 57.4%; Score 58; DB 2; Length 331;
Best Local Similarity 58.3%; Pred. No. 1.98e+00;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

DB 1 MVTYRSAGVDID 12
QY 4 KITRISGVGID 15

RESULT 5
ENTRY T01000
TITLE hypothetical protein T517.1 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change
05-Feb-1999
ACCESSION T01000
REFERENCE Z14159
#authors Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.;
Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.;
Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter,
J.C.
#submission submitted to the EMBL Data Library, November 1997
#description Arabidopsis thaliana chromosome II BAC T517 genomic sequence.
#accession T01000
#molecule_type DNA
#status preliminary; translated from GB/EMBL/DDBT
#residues 1-441 ##label ROU
#cross-references EMBL:AC003000; NID:92642152; PID:92642153

GENETICS
#map_position II
#note T517.1
SUMMARY #length 441 #molecular-weight 47873 #checksum 2617

Query Match 56.4%; Score 57; DB 2; Length 441;
Best Local Similarity 40.0%; Pred. No. 3.07e+00;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

DB 366 SGGKLLRYVNGAGSE 380
QY 1 ATOKITRISGVGID 15

RESULT 6
ENTRY VG1HNM
TITLE E2 glycoprotein precursor - bovine coronavirus (strain Mebus)
ALTERNATE_NAMES peplomer glycoprotein; S glycoprotein; spike glycoprotein
CONTAINS 90A glycoprotein; 90B glycoprotein
ORGANISM #formal_name bovine coronavirus
DATE 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change
20-Mar-1998
ACCESSION A34607
REFERENCE A34607
#authors Abraham, S.; Klenzle, T.E.; Lapps, W.; Brian, D.A.
#journal Virology (1990) 176:296-301
#title Deduced sequence of the bovine coronavirus spike protein and
identification of the internal proteolytic cleavage site.
#cross-references MUID:90232743
#accession A34607
#molecule_type genomic RNA
#residues 1-1363 ##label ABR
#cross-references GB:M31053; NID:9323361; PID:9323362
CLASSIFICATION #superfamily coronavirus E2 glycoprotein
glycoprotein; transmembrane protein
FEATURE
1-17
18-1363 #domain signal sequence #status predicted #label SIG/
18-768 #product E2 glycoprotein #status predicted #label E2G/
769-1363 #product 90A glycoprotein #status predicted #label E2G/
1312-1328 #domain transmembrane #status predicted #label TMA/

59,133,198,359,437,
444,649,676,696,
714,739,788,895,
937,1194,1224,1234,
1253,1267,1288
#binding_site carbohydrate (Asn) (covalent) #status
predicted

SUMMARY #length 1363 #molecular-weight 150809 #checksum 7580

Query Match 56.4%; Score 57; DB 1; Length 1363;
Best Local Similarity 45.5%; Pred. No. 3.07e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 995 NVOYRINGIGV 1005
QY 4 KITRISGVGI 14

RESULT 7
ENTRY S44241 #type complete
TITLE surface protein - human coronavirus
ORGANISM #formal_name human coronavirus
DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
08-Sep-1997

ACCESSIONS S44241
REFERENCE S44240
#authors Kuenkel, F.; Herrier, G.
#submission submitted to the EMBL Data Library, April 1994
#description Sequence and functional analysis of the surface protein of
two human Coronavirus OC43 isolates adapted to growth in
MDCK I and Vero cells.
#accession S44241
#status Preliminary
#molecule_type DNA
#residues 1-1363 #label KUE
#cross-references EMBL:Z32769; NID:9475893; PID:9475894
CLASSIFICATION #superfamily coronavirus E2 glycoprotein
SUMMARY #length 1363 #molecular-weight 150835 #checksum 5716

Query Match 56.4%; Score 57; DB 2; Length 1363;
Best Local Similarity 45.5%; Pred. No. 3.07e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 995 NVOYRINGIGV 1005
QY 4 KITRISGVGI 14

RESULT 8
ENTRY VG1HL9 #type complete
TITLE E2 glycoprotein precursor - bovine coronavirus (strain L9)
ALTERNATE_NAMES peplomer glycoprotein; S glycoprotein; spike glycoprotein
CONTAINS 90A glycoprotein; 90B glycoprotein
ORGNISM #formal_name bovine coronavirus
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
05-Sep-1997

ACCESSIONS A40320
REFERENCE A40320
#authors Zhang, X.; Kousoulas, K.G.; Storz, J.
#journal Virology (1991) 183:397-404
#title Comparison of the nucleotide and deduced amino acid sequences
of the S genes specified by virulent and avirulent strains
of bovine coronaviruses.
#cross-references MUID:91272503
#accession A40320
#molecule_type genomic RNA
#residues 1-1363 #label ZHA
#cross-references GB:M64667; NID:9323355; PID:9323356
CLASSIFICATION #superfamily coronavirus E2 glycoprotein
KEYWORDS glycoprotein; transmembrane protein
FEATURE
1-17 #domain signal sequence #status predicted #label SIG
18-1363 #product E2 glycoprotein #status predicted #label E2G
18-768 #product 90B glycoprotein #status predicted #label E9B

769-1363
1312-1328
59,133,198,359,437,
444,676,696,714,
739,788,895,937,
1194,1224,1234,
1253,1267,1288
#product 90A glycoprotein #status predicted #label E9A\
#domain transmembrane #status predicted #label TMN\
predicted

SUMMARY #length 1363 #molecular-weight 150805 #checksum 5554

Query Match 56.4%; Score 57; DB 1; Length 1363;
Best Local Similarity 45.5%; Pred. No. 3.07e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 995 NVOYRINGIGV 1005
QY 4 KITRISGVGI 14

RESULT 9
ENTRY VG1HOU #type complete
TITLE E2 glycoprotein precursor - bovine coronavirus (strain
Quebec)
ALTERNATE_NAMES peplomer glycoprotein; S glycoprotein; spike glycoprotein
CONTAINS 90A glycoprotein; 90B glycoprotein
ORGANISM #formal_name bovine coronavirus
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
07-Oct-1994

ACCESSIONS A34147
REFERENCE A34147
#authors Parker, M.D.; Yoo, D.; Cox, G.J.; Babluk, L.A.
#journal J. Gen. Virol. (1990) 71:263-270
#title Primary structure of the S peplomer gene of bovine
coronavirus and surface expression in insect cells.
#cross-references MUID:90171910
#accession A34147
#molecule_type genomic RNA
#residues 1-1363 #label PAR
#cross-references GB:D00662
#note In the authors' translation residues 23-31 are shown
after residue 34 and, consequently, residues 32-34 are
displaced nine codons to the left
#note the authors translated the codons UCU, AUU, and AGC for
residues 35-37 as Ile, Ser, and Thr, respectively
the authors translated the codons CCA, GAU, ACU, CCA,
CCU, CCA, GAA, GAU, CCU and GAA for residues 10, 18,
163, 168, 264, 303, 792, 825, 898, and 1125 as Phe,
Arg, Tyr, Phe, Phe, Phe, Gly, Glu, Phe, and Gly,
respectively
#domain signal sequence #status predicted #label SIG\
#product E2 glycoprotein #status predicted #label E2G\
#product 90B glycoprotein #status predicted #label E9B\
#product 90A glycoprotein #status predicted #label E9A\
#domain transmembrane #status predicted #label TMN\
1-17
18-1363
18-768
769-1363
1312-1328
59,133,198,359,437,
649,676,696,714,
739,788,895,937,
1194,1224,1234,
1253,1267,1288
#binding_site carbohydrate (Asn) (covalent) #status
predicted

SUMMARY #length 1363 #molecular-weight 150868 #checksum 6831

Query Match 56.4%; Score 57; DB 1; Length 1363;
Best Local Similarity 45.5%; Pred. No. 3.07e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

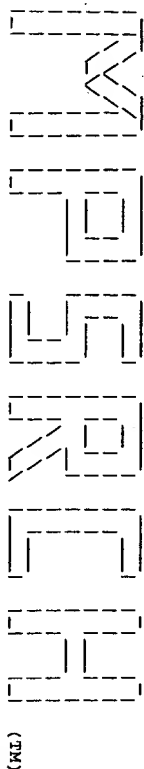
DB 995 NVOYRINGIGV 1005
QY 4 KITRISGVGI 14

RESULT 10
ENTRY
TITLE
ALTERNATE_NAMES
CONTAINS
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MUID:90171938
#accession A34151
#molecule_type genomic RNA
#residues 1-1363 #label BOI
#cross-references GB:D00731; NID:9221136; PID:d1001089; PID:9221137
CLASSIFICATION
#superfamily coronavirus E2 glycoprotein
KEYWORDS
FEATURE
1-17
18-1363
18-768
769-1363
1312-1328
59,133,198,359,437,
444,649,676,696,
714,739,788,895,
937,1194,1224,1234,
1253,1267,1288
#binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY
#length 1363 #molecular-weight 150746 #checksum 5372
Query Match
Best Local Similarity 45.5%; Score 57; DB 1; Length 1363;
Pred. No. 3.07e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 995 NVOYRINGIGV 1005
QY 4 KITRISGVGI 14
RESULT 11
ENTRY
TITLE
ALTERNATE_NAMES
CONTAINS
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MUID:91272503
#accession B40320
#molecule_type genomic RNA
#residues 1-1363 #label ZHA
#cross-references GB:M6468
#superfamily coronavirus E2 glycoprotein
glycoprotein; transmembrane protein
domain signal sequence #status predicted #label SIG\

18-1363
18-768
769-1363
1312-1328
59,133,198,359,437,
444,649,676,696,
714,739,788,895,
937,1194,1224,1234,
1253,1267,1288
#binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY
#length 1363 #molecular-weight 150858 #checksum 7279
Query Match
Best Local Similarity 45.5%; Score 57; DB 1; Length 1363;
Pred. No. 3.07e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 995 NVOYRINGIGV 1005
QY 4 KITRISGVGI 14
RESULT 12
ENTRY
TITLE
ALTERNATE_NAMES
CONTAINS
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MUID:91272503
#accession C40320
#molecule_type genomic RNA
#residues 1-1363 #label ZHA
#cross-references GB:M6469
#superfamily coronavirus E2 glycoprotein
glycoprotein; transmembrane protein
domain signal sequence #status predicted #label SIG\

KEYWORDS glycoprotein; transmembrane protein
FEATURE 1-17
18-1363 #domain signal sequence #status predicted #label SIG
18-1368 #product E2 glycoprotein #status predicted #label E2G
766-1363 #product 90B glycoprotein #status predicted #label EGB
1312-1328 #domain transmembrane #status predicted #label TMN
59,133,198,359,437,
444,649,676,696,
714,739,788,895,
937,1194,1224,1234,
1253,1267,1288 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 1363 #molecular-weight 150781 #checksum 7311
Query Match 56.4%; Score 57; DB 1; Length 1363;
Best Local Similarity 45.5%; Pred. No. 3.07e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
DB 995 NVOYRINGICV 1005
QY 4 KITTYRISGVGI 14
RESULT 14
ENTRY D70163 #type complete
TITLE hypothetical protein B80509 - Lyme disease spirochete
ORGANISM #formal_name Borrelia burgdorferi #common_name Lyme disease
spirochete
DATE 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change
05-Jun-1998
ACCESSIONS D70163
REFERENCE A70100
#authors Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.;
Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.;
Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb,
J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.;
Kerlavange, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
J.; Uitterlinden, T.; Matthey, L.; McDonald, L.; Artlich, P.;
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst,
K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
#journal Nature (1997) 390:580-586
#title Genomic sequence of a Lyme disease spirochete, Borrelia
burgdorferi.
#cross-references M01D:98065943
#accession D70163
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-419 #label KLE
#cross-references GB:AE001153; GB:AE000783; NID:q2688419; PID:q2688429;
TIGR:B80509
SUMMARY #experimental_source strain B31
#length 419 #molecular-weight 48741 #checksum 597
Query Match 55.4%; Score 56; DB 2; Length 419;
Best Local Similarity 38.5%; Pred. No. 4.74e+00;
Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;
DB 278 RNAYRININVD 290
QY 3 QKITRISGVGI 15
RESULT 15
ENTRY J01428 #type complete
TITLE hypothetical 11.6K protein - strawberry mild yellow
edge-associated virus
ALTERNATE_NAMES ORF3 protein
ORGANISM #formal_name strawberry mild yellow edge-associated virus
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
31-Oct-1997

ACCESSIONS J01428
REFERENCE J01426
#authors Jellmann, W.; Mais, E.; Martin, R.R.
#journal J. Gen. Virol. (1992) 73:475-479
#title The nucleotide sequence and genome organization of strawberry
mild yellow edge-associated potexvirus.
#cross-references M01D:92166762
#accession J01428
#molecule_type genomic RNA
#residues 1-108 #label JEL
#cross-references GB:D12517; DDBJ:D01227; NID:q222631; PID:d1002574;
PID:q222634
SUMMARY #experimental_source strain MY-18
#length 108 #molecular-weight 11589 #checksum 628
KEYWORDS transmembrane protein
Query Match 54.5%; Score 55; DB 2; Length 108;
Best Local Similarity 58.3%; Pred. No. 7.28e+00;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
DB 9 HSITRIHAYGL 20
QY 3 QKITRISGVGI 14
Search completed: Fri Jun 11 17:13:37 1999
Job time : 84 secs.



(TM)

Release 3.1A John F. Collins, Bioinformatics Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 11 17:09:17 1999; Maspar time 2.55 Seconds

Tabular output not generated. 166.490 Million cell updates/sec

Title: >US-08-991-628-1
Description: (1-15) from US08991628.pep
Perfect Score: 101
Sequence: 1 ATOKITRISGVGID 15

Scoring table:
PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 26.539; Variance 30.462; scale 0.871

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	101	100.0	999	1	DSG3_HUMAN DESMOGLEIN 3 PRECURSOR	3.98e-12
2	87	86.1	1043	1	DSG1_BOVIN DESMOGLEIN 1 PRECURSOR	3.06e-08
3	87	86.1	1049	1	DSG1_HUMAN DESMOGLEIN 1 PRECURSOR	3.06e-08
4	62	61.4	388	1	PUR5_VIGUN PHOSPHORIBOSYLFORMYLGL	5.81e-02
5	57	56.4	1363	1	VGL2_CVBM E2 GLYCOPROTEIN PRECUR	7.46e-01
6	57	56.4	1363	1	VGL2_CVBLX E2 GLYCOPROTEIN PRECUR	7.46e-01
7	57	56.4	1363	1	VGL2_CVBQ E2 GLYCOPROTEIN PRECUR	7.46e-01
8	57	56.4	1363	1	VGL2_CVBV E2 GLYCOPROTEIN PRECUR	7.46e-01
9	57	56.4	1363	1	VGL2_CVBF E2 GLYCOPROTEIN PRECUR	7.46e-01
10	57	56.4	1363	1	VGL2_CVBL9 E2 GLYCOPROTEIN PRECUR	7.46e-01
11	55	54.5	108	1	MEME_SMYEA 11.5 KD MEMBRANE PROTE	1.99e+00
12	55	54.5	159	1	V459_MYCPN HYPOTHETICAL PROTEIN M	1.99e+00
13	55	54.5	446	1	AP50_SCHPO PUTATIVE CLATHRIN COAT	1.99e+00
14	55	54.5	897	1	BETA-GALACTOSIDASE (EC	1.99e+00
15	54	53.5	479	1	CRU1_RAPSA CRUCIFERIN PCGRSSES P	3.21e+00
16	54	53.5	509	1	CRU3_BRANA CRUCIFERIN CRU1 PRECUR	3.21e+00
17	54	53.5	531	1	GSPE_XANCP GENERAL SECRETION PATH	3.21e+00
18	54	53.5	884	1	CAD1_HUMAN EPITHELIAL-CADHERIN PR	3.21e+00
19	54	53.5	884	1	CAD1_MOUSE EPITHELIAL-CADHERIN PR	3.21e+00
20	53	52.5	343	1	YDE3_SCHPO HYPOTHETICAL 61.8 KD P	5.14e+00
21	53	52.5	890	1	BCN5_CLOPE BACTERIOCIDIN BCN5.	5.14e+00
22	53	52.5	1235	1	VGL2_CVMJH E2 GLYCOPROTEIN PRECUR	5.14e+00
23	53	52.5	1353	1	VGL2_CVHOC E2 GLYCOPROTEIN PRECUR	5.14e+00

24	53	52.5	1371	1	PUR2_CHITE PHOSPHORIBOSYLAMINE--G	5.14e+00
25	53	52.5	1376	1	VGL2_CVMJC E2 GLYCOPROTEIN PRECUR	5.14e+00
26	53	52.5	1376	1	VGL2_CVM4 E2 GLYCOPROTEIN PRECUR	5.14e+00
27	52	51.5	243	1	TONE_KLEPN TONOR PROTEIN.	8.19e+00
28	52	51.5	435	1	HFE1_HAELN MINOR FIBRININ SUBUNIT	8.19e+00
29	52	51.5	956	1	TSP3_MOUSE THROMBOSPONDIN 3 PREC	8.19e+00
30	52	51.5	1121	1	YJ09_YEAST HYPOTHETICAL 127.4 KD	8.19e+00
31	52	51.5	1319	1	DYNA_DROME 150 KD DYNEIN-ASSOCIAT	8.19e+00
32	52	51.5	1324	1	VGL2_CVMA5 E2 GLYCOPROTEIN PRECUR	8.19e+00
33	51	50.5	68	1	S3AC_BACSU STAGE III SPORULATION	1.30e+01
34	51	50.5	465	1	YB57_YEAST HYPOTHETICAL 51.5 KD P	1.30e+01
35	51	50.5	493	1	IMDH_STEYP INOSINE-5'-MONOPHOSPHA	1.30e+01
36	51	50.5	504	1	MYOC_HUMAN MYOCILIN PRECURSOR (TR	1.30e+01
37	51	50.5	513	1	IMDH_BACSU INOSINE-5'-MONOPHOSPHA	1.30e+01
38	51	50.5	887	1	CAD1_CHICK EPITHELIAL-CADHERIN PR	1.30e+01
39	51	50.5	913	1	CAD4_MOUSE RETINAL-CADHERIN PREC	1.30e+01
40	51	50.5	916	1	CAD4_HUMAN RETINAL-CADHERIN PREC	1.30e+01
41	51	50.5	1218	1	YK83_YEAST PROBABLE ATP-DEPENDENT	1.30e+01
42	51	50.5	1271	1	BCR_HUMAN BREAKPOINT CLUSTER REG	1.30e+01
43	51	50.5	1709	1	CHD1_HUMAN CHROMODOMAIN-HELICASE-	1.30e+01
44	51	50.5	1711	1	CHD1_MOUSE CHROMODOMAIN-HELICASE-	1.30e+01
45	51	50.5	1902	1	P2P_LACPA P11-TYPE PROTEINASE PR	1.30e+01

ALIGNMENTS

RESULT ID	1	STANDARD;	PRT;	999 AA.
AC	DSG3_HUMAN			
AC	P32926;			
DT	01-OCT-1993 (REL. 27, CREATED)			
DT	01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	DESMOGLIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).			
GN	DSG3.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;			
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 92069753.			
RA	AMAGAI M., KLAUS-KOYTUN V., STANLEY J.R.;			
RT	"Autoantibodies against a novel epithelial cadherin in pemphigus			
RT	vulgaris, a disease of cell adhesion.";			
RL	CELL 67:869-877(1991).			
CC	-1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.			
CC	INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE			
CC	FILAMENT'S MEDIATING CELL-CELL ADHESION.			
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.			
CC	-1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, OESOPHAGUS AND			
CC	CARCINOMAS.			
CC	-1- DOMAIN: CALCITONIN MAY BE BOUND BY THE CADHERIN-LIKE REPEATS			
CC	(POTENTIAL).			
CC	-1- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN			
CC	DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE			
CC	LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES			
CC	AGAINST DSG3.			
CC	-1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE			
CC	DESMOSOMAL SUBFAMILY.			
CC	THIS SWISS-PROT ENTRY IS COPYRIGHTED. IT IS PRODUCED THROUGH A COLLABORATION			
CC	BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -			
CC	THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS			
CC	USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY			
CC	MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL			
CC	ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/			
CC	OR SEND AN EMAIL TO license@isb-sib.ch).			
DR	EMBL; M76482; G190752; -			
DR	PIR; A41088; ICH053.			
DR	MIM; 169615; -			
DR	PROSITE; PS00232; CADHERIN; 3.			
DR	PFAM; PF00028; cadherin; 4.			
DR	HSSP; P09803; 1EDH.			

CELL ADHESION: SIGNAL: TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
CALCIUM-BINDING; REPEAT.
FT SIGNAL 1 23
FT PROPEP 24 49
FT CHAIN 50 999
FT DOMAIN 50 615
FT TRANSMEM 616 640
FT DOMAIN 641 999
FT REPEAT 50 158
FT REPEAT 159 268
FT REPEAT 269 383
FT REPEAT 386 499
FT REPEAT 910 935
FT REPEAT 936 966
FT CARBOHYD 110 110
FT CARBOHYD 180 180
FT CARBOHYD 459 459
FT CARBOHYD 545 545
SQ SEQUENCE 999 AA: 107503 MW: 489166AE CRC32;
Query Match 100.0%; Score 101; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 3,98e-12;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 79 ATOKITRISGVGID 93
OY 1 ATOKITRISGVGID 15
RESULT 2
ID DSG1 BOVIN STANDARD; PRT: 1043 AA.
AC 003763; (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DG1).
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDEA; BOVINA; BOS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-MUZZLE EPITHELIUM;
RA KOCH P.J., GOLDSCHMIDT M.D., ZIMBELMANN R., FRANK W.W.;
RL SUBMITTED (MAR-1991) TO EMBL/GENBANK/DBD DATA BANKS.
RN [2]
RP SEQUENCE OF 44-1043 FROM N.A.
RC TISSUE-MUZZLE EPITHELIUM;
RA KOCH P.J., GOLDSCHMIDT M.D., ZIMBELMANN R., FRANK W.W.;
RX MEDLINE: 91168965.
RA KOCH P.J., WALSH M.J., SCHMELZ M., GOLDSCHMIDT M.D.,
RA ZIMBELMANN R., FRANK W.W.;
RT "Identification of desmoglein, a constitutive desmosomal
glycoprotein, as a member of the cadherin family of cell adhesion
molecules."
RL EUR. J. CELL BIOL. 53:1-12(1990).
RN [3]
RP REVISIONS, AND SEQUENCE OF 101-123.
RX MEDLINE: 92037656.
RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R.,
RA FRANK W.W.;
RT "Complete amino acid sequence of the epidermal desmoglein precursor
polypeptide and identification of a second type of desmoglein gene."
RL EUR. J. CELL BIOL. 55:200-208(1991).
RN [4]
RP SEQUENCE OF 44-493 FROM N.A.
RX MEDLINE: 91097553.
RA GOODMAN L., HILL J.E., RAYNOR K., RASZI L., MANABE M., COWIN P.;
RT "Desmoglein shows extensive homology to the cadherin family of cell
adhesion molecules."
RL BIOCHEM. BIOPHYS. RES. COMMUN. 173:1224-1230(1990).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME
JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
FILAMENTS MEDIATING CELL-CELL ADHESION.

-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EPIDERMIS, MUZZLE, TONGUE AND OESOPHAGUS.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
(POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE
DESMOSOMAL SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: X58466; G307; -
DR EMBL: X57784; G436062; -
DR EMBL: M58165; G552318; -
DR PIR: S14603; IJBOG1.
DR PROSITE: PS00232; CADHERIN; 2.
DR PFAM: PF00028; cadherin; 3.
DR HSSP: P09803; 1EDH.
DR
DR CELL ADHESION: SIGNAL: TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
CALCIUM-BINDING; REPEAT.
KM
FT SIGNAL 1 23
FT PROPEP 24 49
FT CHAIN 50 1043
FT DOMAIN 50 548
FT TRANSMEM 549 573
FT DOMAIN 574 1043
FT REPEAT 50 158
FT REPEAT 159 270
FT REPEAT 271 385
FT REPEAT 386 498
FT REPEAT 499 845
FT REPEAT 846 875
FT REPEAT 876 905
FT REPEAT 906 933
FT REPEAT 934 962
FT DOMAIN 963 1012
FT CARBOHYD 110 110
FT CARBOHYD 180 180
FT CARBOHYD 496 496
FT CONFLICT 124 124
SQ SEQUENCE 1043 AA: 112243 MW: 13898584 CRC32;
Query Match 86.1%; Score 87; DB 1; Length 1043;
Best Local Similarity 80.0%; Pred. No. 3.06e-08;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
DB 79 ANOQVYRISGVGID 93
OY 1 ANOQVYRISGVGID 15
RESULT 3
ID DSG1 HUMAN STANDARD; PRT: 1049 AA.
AC 002413;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DG1).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CARAPRIMIN; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-KERATINOCYTES;
RX MEDLINE: 91271279.
RA WHEELER G.N., PARKER A.E., THOMAS C.L., ATALLOTIS P., POYNTER D.,
RA ARBEMANN J., RUTMAN A.J., PIDSLEY S.C., WAIT F.M., REES D.A.,
RA BUXTON R.S., MAGEE A.I.;

RT "Desmosomal glycoprotein Dsg, a component of intercellular desmosome
RT junctions, is related to the cadherin family of cell adhesion
RT molecules.";
RT PROC. NATL. ACAD. SCI. U.S.A. 88:4796-4800(1991).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLACUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSILAND OESOPHAGUS.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE
CC DESMOsome SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X56554; G30506; .
DR PIR; S16906; IJHUG1.
DR MIM; 125670; .
DR PROSITE; PS00232; CADHERIN; 2.
DR PFM; PF00028; cadherin; 4.
DR HSSP; P09803; 1EDH.
KW CELL ADHESION; SIGNAL; TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
KW CALCIUM-BINDING; REPEAT.
FT SIGNAL 1 23
FT PROPEP 24 49 POTENTIAL.
FT CHAIN 50 1049 DESMOGLEIN 1.
FT DOMAIN 50 545 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 546 570 POTENTIAL.
FT DOMAIN 571 1049 CYTOPLASMIC (POTENTIAL).
FT REPEAT 50 158 CADHERIN 1.
FT REPEAT 159 270 CADHERIN 2.
FT REPEAT 271 385 CADHERIN 3.
FT REPEAT 386 497 CADHERIN 4.
FT REPEAT 497 839 DESMOGLEIN REPEAT 1.
FT REPEAT 839 869 DESMOGLEIN REPEAT 2.
FT REPEAT 869 899 DESMOGLEIN REPEAT 3.
FT REPEAT 899 927 DESMOGLEIN REPEAT 4.
FT REPEAT 927 956 DESMOGLEIN REPEAT 5.
FT DOMAIN 956 1019 GLY/SER-RICH.
FT CARBOHYD 36 36 POTENTIAL.
FT CARBOHYD 110 110 POTENTIAL.
FT CARBOHYD 180 180 POTENTIAL.
SQ SEQUENCE 1049 AA; FDD/9961 CRC32;
Query Match 86.1%; Score 87; DB 1; Length 1049;
Best Local Similarity 80.0%; Pred. No. 3,06e-08;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DB 79 ANQVTRISGVGID 93
OY 1 ATOKITRISGVGID 15
RESULT 4
ID PURS-VIGUN STANDARD; PRT; 388 AA.
AC P52424;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PHOSPHORIBOSYL-FORMYLGLYCINAMIDINE CYCLO-LIGASE PRECURSOR (EC 6.3.3.1)
DE (AIRS) (PHOSPHORIBOSYL-AMINOIMIDAZOLE SYNTHETASE) (AIR SYNTHASE).
GN PURS.
OS VIGNA UNGUICULATA (COMPEA).
OC EUKARYOTA; VIRIDIPHYTA; STREPTOPHYTA; EMEROPHYTA; TRACHEOPHYTA;
OC EUKARYOTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC FABACEAE; PAPILIONOIDEAE; VIGNA.

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. VITA 3; TISSUE-ROOT NODULES;
RA SMITH P.M.C., MANN A.J., HALL D.J., ATKINS C.A.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DDAT DATA BANKS.
CC -1- CATALYTIC ACTIVITY: ATP + 5'-PHOSPHORIBOSYL-FORMYLGLYCINAMIDINE -
CC ADP + ORTHOPHOSPHATE + 5'-PHOSPHORIBOSYL-5-AMINOIMIDAZOLE.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -1- SIMILARITY: TO OTHER AIRS FROM BACTERIA AND EUKARYOTES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U30895; G945050; .
DR PFM; PF00586; AIRS; 1.
DR PURINE BIOSYNTHESIS; LIGASE; CHLOROPLAST; TRANSIT PEPTIDE.
KW TRANSIT 1 388
KW CHAIN 1 388
FT CHAIN 1 388
SQ SEQUENCE 388 AA; 40427 MW; 895D3159 CRC32;
Query Match 61.4%; Score 62; DB 1; Length 388;
Best Local Similarity 60.0%; Pred. No. 5,81e-02;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

DB 62 ASQGLTYRDAGYDID 76
OY 1 ATOKITRISGVGID 15
RESULT 5
ID VGL2-CVBM STANDARD; PRT; 1363 AA.
AC P15777;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-JUN-1999 (REL. 29, LAST ANNOTATION UPDATE)
DE E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOWER PROTEIN).
GN S.
OS BOVINE CORONAVIRUS (STRAIN MEUS).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; NIDOVIRALES;
OC CORONAVIRIDAE; CORONAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE; 90232743.
RA ABRAHAM S., KIENZLE T.E., LAPP S.W.E., BRIAN D.A.;
RT "Deduced sequence of the bovine coronavirus spike protein and
RT identification of the internal proteolytic cleavage site.";
RL VIROLOGY 176:296-301(1990).
RN [2]
RP SEQUENCE OF 1254-1363 FROM N.A.
RX MEDLINE; 90320120.
RA ABRAHAM S., KIENZLE T.E., LAPP S.W.E., BRIAN D.A.;
RT "Sequence and expression analysis of potential nonstructural proteins
RT of 4.9, 4.8, 12.7, and 9.5 kDa encoded between the spike and membrane
RT protein genes of the bovine coronavirus.";
RL VIROLOGY 177:488-495(1990).
CC -1- FUNCTION: THE PEPLOWER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----


```

FT CHAIN 769 1363 SPIKE PROTEIN S2 (90A).
FT DOMAIN 18 1311 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1312 1328 POTENTIAL.
FT DOMAIN 1329 1363 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1329 1346 CYS-RICH.
FT CARBOHYD 59 59 POTENTIAL.
FT CARBOHYD 133 133 POTENTIAL.
FT CARBOHYD 198 198 POTENTIAL.
FT CARBOHYD 359 359 POTENTIAL.
FT CARBOHYD 437 437 POTENTIAL.
FT CARBOHYD 649 649 POTENTIAL.
FT CARBOHYD 676 676 POTENTIAL.
FT CARBOHYD 696 696 POTENTIAL.
FT CARBOHYD 714 714 POTENTIAL.
FT CARBOHYD 739 739 POTENTIAL.
FT CARBOHYD 788 788 POTENTIAL.
FT CARBOHYD 937 937 POTENTIAL.
FT CARBOHYD 1194 1194 POTENTIAL.
FT CARBOHYD 1224 1224 POTENTIAL.
FT CARBOHYD 1234 1234 POTENTIAL.
FT CARBOHYD 1253 1253 POTENTIAL.
FT CARBOHYD 1267 1267 POTENTIAL.
FT CARBOHYD 1288 1288 POTENTIAL.
SQ SEQUENCE 1363 AA; 150869 MW; CB76E91A CRC32;

Query Match 56.4%; Score 57; DB 1; Length 1363;
Best Local Similarity 45.5%; Pred. No. 7.46e-01;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 995 NVOYRINGIGV 1005
Oy 4 KITRISGVGI 14

RESULT 8 STANDARD; PRT; 1363 AA.
ID VGL2_CVBF
AC P25190;
BT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DE 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOMER PROTEIN).
GN S.
OS BOVINE CORONAVIRUS (STRAIN F15).
OC VIRUSES; SSRA POSITIVE-STRAND VIRUSES, NO DNA STAGE; NIDOVIRALES;
OC CORONAVIRIDAE; CORONAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91272503.
RA ZHANG X., KOUSOULOS K.G., STORZ J.;
RT "Comparison of the nucleotide and deduced amino acid sequences of the
RT S genes specified by virulent and avirulent strains of bovine
RT coronaviruses."
RL VIROLOGY 183:397-404(1991).
CC -1- FUNCTION: THE PELOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M64668; G454810; -
DR PIR; B40320; VGTHVA.
KW GLYCOPROTEIN; ENVELOPE PROTEIN; TRANSMEMBRANE; SIGNAL.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 1363 SPIKE E2 GLYCOPROTEIN.
FT CHAIN 18 768 SPIKE PROTEIN S1 (90B).
FT CHAIN 18 768 SPIKE PROTEIN S2 (90A).
FT DOMAIN 18 1311 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1312 1328 POTENTIAL.

```

```

FT DOMAIN 1329 1363 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1329 1346 CYS-RICH.
FT CARBOHYD 59 59 POTENTIAL.
FT CARBOHYD 133 133 POTENTIAL.
FT CARBOHYD 198 198 POTENTIAL.
FT CARBOHYD 359 359 POTENTIAL.
FT CARBOHYD 437 437 POTENTIAL.
FT CARBOHYD 649 649 POTENTIAL.
FT CARBOHYD 676 676 POTENTIAL.
FT CARBOHYD 696 696 POTENTIAL.
FT CARBOHYD 714 714 POTENTIAL.
FT CARBOHYD 739 739 POTENTIAL.
FT CARBOHYD 788 788 POTENTIAL.
FT CARBOHYD 937 937 POTENTIAL.
FT CARBOHYD 1194 1194 POTENTIAL.
FT CARBOHYD 1224 1224 POTENTIAL.
FT CARBOHYD 1234 1234 POTENTIAL.
FT CARBOHYD 1253 1253 POTENTIAL.
FT CARBOHYD 1267 1267 POTENTIAL.
FT CARBOHYD 1288 1288 POTENTIAL.
SQ SEQUENCE 1363 AA; 150843 MW; E0444ED6 CRC32;

Query Match 56.4%; Score 57; DB 1; Length 1363;
Best Local Similarity 45.5%; Pred. No. 7.46e-01;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 995 NVOYRINGIGV 1005
Oy 4 KITRISGVGI 14

RESULT 9 STANDARD; PRT; 1363 AA.
ID VGL2_CVBF
AC P25190;
BT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DE 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOMER PROTEIN).
GN S.
OS BOVINE CORONAVIRUS (STRAIN F15).
OC VIRUSES; SSRA POSITIVE-STRAND VIRUSES, NO DNA STAGE; NIDOVIRALES;
OC CORONAVIRIDAE; CORONAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90171938.
RA BOIREAU P., CRUCIERE C., LAPORTE J.;
RT "Nucleotide sequence of the glycoprotein S gene of bovine enteric
RT coronavirus and comparison with the S proteins of two mouse hepatitis
RT virus strains."
RL J. GEN. VIROL. 71:487-492(1990).
CC -1- FUNCTION: THE PELOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; D00731; G221137; -
DR PIR; A34151; VGTHFL.
KW GLYCOPROTEIN; ENVELOPE PROTEIN; TRANSMEMBRANE; SIGNAL.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 1363 SPIKE E2 GLYCOPROTEIN.
FT CHAIN 18 768 SPIKE PROTEIN S1 (90B).
FT CHAIN 18 768 SPIKE PROTEIN S2 (90A).
FT TRANSMEM 1312 1328 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1329 1363 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1329 1346 CYS-RICH.
FT CARBOHYD 59 59 POTENTIAL.

```

```

FT CARBOHYD 133 133 POTENTIAL.
FT CARBOHYD 198 198 POTENTIAL.
FT CARBOHYD 359 359 POTENTIAL.
FT CARBOHYD 437 437 POTENTIAL.
FT CARBOHYD 649 649 POTENTIAL.
FT CARBOHYD 676 676 POTENTIAL.
FT CARBOHYD 696 696 POTENTIAL.
FT CARBOHYD 714 714 POTENTIAL.
FT CARBOHYD 739 739 POTENTIAL.
FT CARBOHYD 788 788 POTENTIAL.
FT CARBOHYD 937 937 POTENTIAL.
FT CARBOHYD 1194 1194 POTENTIAL.
FT CARBOHYD 1224 1224 POTENTIAL.
FT CARBOHYD 1234 1234 POTENTIAL.
FT CARBOHYD 1253 1253 POTENTIAL.
FT CARBOHYD 1267 1267 POTENTIAL.
FT CARBOHYD 1288 1288 POTENTIAL.
SQ SEQUENCE 1363 AA; 150746 MW; ABCD5933 CRC32;

Query Match
Best Local Similarity 45.5%; Score 57; DB 1; Length 1363;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 995 NVOYRISGV 1005
QY 4 KITYRISGVGI 14

RESULT 10
ID VGL2_CVBL9 STANDARD; PRT; 1363 AA.
AC P25191;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DE 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOMER PROTEIN).
GN S.
OS BOVINE CORONAVIRUS (STRAIN L9).
OC VIRUSES; SSNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; NIDOVIRALES;
OC CORONAVIRIDAE; CORONAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91272503.
RA ZHANG X., KOUSOULAS K.G., STORZ J.;
RT "Comparison of the nucleotide and deduced amino acid sequences of the
S genes specified by virulent and avirulent strains of bovine
coronaviruses";
RT VIROLOGY 183:397-404(1991).
CC -1- FUNCTION: THE PELOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; M64667; G323356;
DR PIR; A40320; VGIH19.
KW GLYCOPROTEIN; ENVELOPE PROTEIN; TRANSMEMBRANE; SIGNAL.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 1363 SPIKE E2 GLYCOPROTEIN.
FT CHAIN 18 768 SPIKE PROTEIN S1 (90B).
FT CHAIN 769 1363 SPIKE PROTEIN S2 (90A).
FT DOMAIN 18 1311 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1312 1328 POTENTIAL.
FT DOMAIN 1329 1363 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1329 1346 CYS-RICH.
FT CARBOHYD 59 133 POTENTIAL.
FT CARBOHYD 133 133 POTENTIAL.
FT CARBOHYD 198 198 POTENTIAL.
FT CARBOHYD 359 359 POTENTIAL.

```

```

FT CARBOHYD 437 437 POTENTIAL.
FT CARBOHYD 676 676 POTENTIAL.
FT CARBOHYD 696 696 POTENTIAL.
FT CARBOHYD 714 714 POTENTIAL.
FT CARBOHYD 739 739 POTENTIAL.
FT CARBOHYD 788 788 POTENTIAL.
FT CARBOHYD 937 937 POTENTIAL.
FT CARBOHYD 1194 1194 POTENTIAL.
FT CARBOHYD 1224 1224 POTENTIAL.
FT CARBOHYD 1234 1234 POTENTIAL.
FT CARBOHYD 1253 1253 POTENTIAL.
FT CARBOHYD 1267 1267 POTENTIAL.
FT CARBOHYD 1288 1288 POTENTIAL.
SQ SEQUENCE 1363 AA; 150806 MW; 54233485 CRC32;

Query Match
Best Local Similarity 45.5%; Score 57; DB 1; Length 1363;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 995 NVOYRISGV 1005
QY 4 KITYRISGVGI 14

RESULT 11
ID VDEM_SMYEA STANDARD; PRT; 108 AA.
AC 000842;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE 11.5 KD MEMBRANE PROTEIN (ORE 3).
OS STRAWBERRY MILD YELLOW EDGE-ASSOCIATED VIRUS (SMVEAV).
OC VIRUSES; SSNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTEXVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MY-18;
RX MEDLINE; 9216762.
RA JELKMAN W., MAISS E., MARTIN R.R.;
RT "The nucleotide sequence and genome organization of strawberry mild
yellow edge-associated potexvirus";
RT J. GEN. VIROL. 73:475-479(1992).
RL -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO ORE3 PROTEIN FROM OTHER POTEXVIRUSES AND TO 12 KD
PROTEIN FROM CARLAVIRUSES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; D12517; G222634;
DR PIR; JQ1428; JQ1428.
DR PIR; PFO1307; Plant_vir_prot; 1.
KW TRANSMEMBRANE.
FT TRANSMEM 15 31 POTENTIAL.
FT TRANSMEM 68 85 POTENTIAL.
FT TRANSMEM 108 AA; 51DFAF03 CRC32;
SQ SEQUENCE 108 AA; 11589 MW; 51DFAF03 CRC32;

Query Match
Best Local Similarity 58.3%; Score 55; DB 1; Length 108;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 9 HSTYRILAVGL 20
QY 3 QKITRISGVGI 14

RESULT 12
ID Y459_MYCPN STANDARD; PRT; 169 AA.
AC P75118;

```

DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE MYOCELLULAR PROTEIN MG459 HOMOLOG.
OS MYOCELLULAR PNEUMONIAE.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLICUTES;
OC MYOPLASMA/ACETABACULUS; MYOPLASMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE; 91105885.
RA HIMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,
RA HERMANN R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
pneumoniae."
RL NUCLEIC ACIDS RES. 24:4420-4449(1996).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000018; G1673831; -
KM HYPOTHETICAL PROTEIN.
SQ SEQUENCE 169 AA; 19466 MW; 86ED0299 CRC32;
Query Match 54.5%; Score 55; DB 1; Length 169;
Best Local Similarity 46.7%; Pred. No. 1.99e+00;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
DB 18 AEOKNRFGVIGVEIE 32
QY 1 ATQKITRISGVGID 15
RESULT 13
ID AP50_SCHPO STANDARD; PRT; 446 AA.
AC 009718;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE PUTATIVE CLATHRIN COAT ASSEMBLY PROTEIN AP50 (CLATHRIN COAT ASSOCIATED
PROTEIN AP50) (PLASMA MEMBRANE ADAPTOR AP-2 50 KD PROTEIN) (HA2 50 KD
SUBUNIT) (CLATHRIN ASSEMBLY PROTEIN COMPLEX 2 MEDIUM CHAIN).
DE SPAC31A2.09C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA DEVIN K., CHURCHER C.M., BARRELL B.G., RAJANDRAM M.A., WALSH S.V.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: COMPONENT OF THE ADAPTOR COMPLEXES WHICH LINK CLATHRIN
TO RECEPTORS IN COATED VESICLES. CLATHRIN-ASSOCIATED PROTEIN
COMPLEXES ARE BELIEVED TO INTERACT WITH THE CYTOPLASMIC TAILS OF
MEMBRANE PROTEINS, LEADING TO THEIR SELECTION AND CONCENTRATION.
CC AP50 IS A SUBUNIT OF THE PLASMA MEMBRANE ADAPTOR (POTENTIAL).
CC -1- SUBUNIT: ASSEMBLY PROTEIN COMPLEX 2 (AP-2) IS AN HETEROTETRAMER
COMPOSED OF TWO LARGE CHAINS (ALPHA AND BETA), A MEDIUM CHAIN
(AP50) AND A SMALL CHAIN (AP17) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: COMPONENT OF THE COAT SURROUNDING THE
CYTOPLASMIC FACE OF COATED VESICLES IN THE PLASMA MEMBRANE
(BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ADAPTOR COMPLEXES MEDIUM SUBUNITS
FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; 250113; G914887; -
DR PROSITE; P500980; CLAT_ADAPTOR_M.1; 1.
DR PROSITE; P500991; CLAT_ADAPTOR_M.2; 1.
DR PFAM; PF00928; Adap_comp_sub; 1.
KM HYPOTHETICAL PROTEIN; COATED PTS.
SQ SEQUENCE 446 AA; 50821 MW; C97D830C CRC32;
Query Match 54.5%; Score 55; DB 1; Length 446;
Best Local Similarity 70.0%; Pred. No. 1.99e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
DB 310 SKOKITRIS 319
QY 1 ATQKITRIS 10
RESULT 14
ID BGAL_CLOBB STANDARD; PRT; 897 AA.
AC P24131;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE).
GN CBGA.
OS CLOSTRIDIUM ACETOBUTYLICUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCIB 2951;
RX MEDLINE; 91216979.
RA HANCOCK K.R., ROCKMAN E., YOUNG C.A., PEARCE L., MADDOX I.S.,
RA SCOTT D.B.;
RT "Expression and nucleotide sequence of the Clostridium acetobutylicum
beta-galactosidase gene cloned in Escherichia coli."
RL J. BACTERIOL. 173:3084-3095(1991).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -1- INDUCTION: LATE IN THE ABE (ACETONE, BUTANOL, AND ETHANOL)
FERMENTATION AND SUBJECT TO GLUCOSE REPRESSION.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M35107; G144746; -
DR PIR; A39405; A39405.
DR PROSITE; P500719; GLYCOSYL_HYDROL_F2.1; 1.
DR PROSITE; P500608; GLYCOSYL_HYDROL_F2.2; 1.
DR PFAM; PF00703; glycosyl_hydrof; 1.
DR HSSP; P00722; 1BGL.
KM HYDROLASE; GLYCOSIDASE.
FT ACT SITE 459 459
FT ACT SITE 503 503
SQ SEQUENCE 897 AA; 105020 MW; 5CFE208B CRC32;
Query Match 54.5%; Score 55; DB 1; Length 897;
Best Local Similarity 72.7%; Pred. No. 1.99e+00;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
DB 828 KITVEVSGEET 838
|||||:||||

QY 4 KITIRISGVCI 14

```

RESULT 15
ID CRUL RAPS A STANDARD; PRT; 479 AA.
AC 00248;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CRUCIFERIN PGCRURSES PRECURSOR (11S GLOBULIN) (12S STORAGE PROTEIN).
GN CRUR.
OS RAPHANUS SATIVUS (RADISH).
OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLIOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPRIFALES; BRASSICACEAE; RAPHANUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. SAKA KNACKER;
RX MEDLINE: 93043037.
RA DEPIGNY-THIS D., RAYNAL M., ASPART L., DELESENY M., GRELLET F.;
RT "The cruciferin gene family in radish.";
RL PLANT MOL. BIOL. 20:467-479(1992)
CC -1- FUNCTION: THIS IS A SEED STORAGE PROTEIN.
CC -1- SUBUNIT: HEXAMER; EACH SUBUNIT IS COMPOSED OF AN ACIDIC AND A
CC BASIC CHAIN DERIVED FROM A SINGLE PRECURSOR AND LINKED BY A
CC DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE 11S SEED STORAGE PROTEINS (GLOBULINS)
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X59808; G21118; -.
DR PIR: S26223; S26223.
DR PROSITE: PS00305; 11S_SEED_STORAGE; 1.
DR PFAM: PF00190; Seedstore_11s; 1.
KW SEED; STORAGE PROTEIN; SIGNAL; MULTIGENE FAMILY.
FT SIGNAL 1 23 BY SIMILARITY.
FT CHAIN 24 289 CRUCIFERIN PGCRURSES ALPHA CHAIN.
FT CHAIN 290 479 CRUCIFERIN PGCRURSES BETA CHAIN.
FT DISULFID 113 296 CRUCIFERIN PGCRURSES BETA CHAIN.
FT DOMAIN 121 141 INTERCHAIN (ALPHA-BETA) (POTENTIAL).
FT DOMAIN 211 218 GLN/GLY-RICH.
FT DOMAIN POLY-GLN.
SQ SEQUENCE 479 AA; 53256 MW; 2A5DEA90 CRC32;

```

Query Match 53.5%; Score 54; DB 1; Length 479;

Best Local Similarity 42.9%; Pred. No. 3,21e+00; Mismatches 3; Indels 0; Gaps 0;

DB 92 SSPKIAVVOGMCI 105

QY 1 ATOKITRISGVCI 14

Search completed: Fri Jun 11 17:09:28 1999
Job time : 11 secs.

(TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_bp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 11 17:09:46 1999; Maspar time 6.16 Seconds

Tabular output not generated. 132,927 Million cell updates/sec

Title: >US-08-991-628-1
Description: (1-15) from US08991628.pep
Perfect Score: 101
Sequence: 1 ATOKITRISGVGID 15

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mmc 8:sp-organelle
9:sp-phase 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 25.479; Variance 31.590; scale 0.807

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	92	91.1	993	11	035902	DESMOGLEIN 3 (FRAGMENT	8.44e+09
2	60	59.4	3380	5	024292	ADHERIN.	3.90e-01
3	59	58.4	2269	5	077360	MAL3P4.25 PROTEIN.	6.33e+00
4	58	57.4	331	2	066968	PHOSPHORIBOSYLFORMYLGL	1.03e+00
5	57	56.4	441	10	022282	TS17.1 PROTEIN.	1.66e+00
6	57	56.4	1363	14	066290	SURFACE PROTEIN PRECUR	1.66e+00
7	56	55.4	1363	14	066290	SURFACE PROTEIN PRECUR	1.66e+00
8	56	55.4	419	2	051462	HYPOTHEICAL 48.7 KD P	2.67e+00
9	55	54.5	108	14	041278	STRAWBERRY MILD YELLOW	4.25e+00
10	54	53.5	287	5	020481	FA6G10.3 PROTEIN.	6.74e+00
11	54	53.5	482	2	059326	CHITINASE PRECURSOR (E	6.74e+00
12	54	53.5	878	4	015855	HYOMOUTIN PRECURSOR (6.74e+00
13	54	53.5	2610	5	019482	F15B9.7 PROTEIN.	6.74e+00
14	53	52.5	82	7	077862	MHC CLASS II B LOCUS 4	1.06e+01
15	53	52.5	82	7	031536	MHC CLASS II BETA CHAI	1.06e+01
16	53	52.5	82	7	031537	MHC CLASS II BETA CHAI	1.06e+01
17	53	52.5	82	7	031538	MHC CLASS II BETA CHAI	1.06e+01
18	53	52.5	280	3	005871	CHROMOSOME XII COSMID	1.06e+01
19	53	52.5	420	3	068332	HYPOTHEICAL 43.9 KD P	1.06e+01
20	53	52.5	780	2	P77846	CELODEXTRIN-PHOSPHORY	1.06e+01

21	53	52.5	832	4	013693	ABR.	1.06e+01
22	53	52.5	859	4	012879	GUANINE NUCLEOTIDE RES	1.06e+01
23	53	52.5	940	10	023327	HYPOTHEICAL 108.0 KD	1.06e+01
24	53	52.5	1177	10	023377	PDR5-LIKE ABC TRANSPOR	1.06e+01
25	53	52.5	1361	14	066199	SURFACE PROTEIN.	1.06e+01
26	53	52.5	1361	14	086623	SURFACE GLYCOPROTEIN S	1.06e+01
27	53	52.5	1376	14	083330	S GLYCOPROTEIN.	1.06e+01
28	52	51.5	154	10	041913	12S STORAGE PROTEIN (F	1.66e+01
29	52	51.5	177	3	012036	P2604 PROTEIN (ORF YPL	1.66e+01
30	52	51.5	301	5	020053	F35G12.10 PROTEIN.	1.66e+01
31	52	51.5	303	2	P95639	HYDROXYMETHYLGLUTARYL-	1.66e+01
32	52	51.5	334	2	P72978	C-TYPE CYTOCHROME SYNT	1.66e+01
33	52	51.5	334	2	055316	HYPOTHEICAL 36.1 KD P	1.66e+01
34	52	51.5	342	2	055316	CAP5E.	1.66e+01
35	52	51.5	342	2	P95699	CAP8E.	1.66e+01
36	52	51.5	524	10	096318	12S CRUCIFERIN SEED ST	1.66e+01
37	52	51.5	572	5	023919	PHOSPHOGLUCOMUTASE A.	1.66e+01
38	52	51.5	575	2	053614	ABCA (FRAGMENT).	1.66e+01
39	52	51.5	765	5	045585	F56H6.7 PROTEIN.	1.66e+01
40	52	51.5	922	13	P79883	XMN-ADHERIN (FRAGMENT	1.66e+01
41	52	51.5	947	5	001318	F02E9.6 PROTEIN.	1.66e+01
42	52	51.5	1334	14	039227	E2 GLYCOPROTEIN PRECUR	1.66e+01
43	52	51.5	1360	14	055253	SPIKE GLYCOPROTEIN.	1.66e+01
44	52	51.5	1361	14	P90210	SPIKE PROTEIN.	1.66e+01
45	52	51.5	1361	14	Q83331	S GLYCOPROTEIN.	1.66e+01

ALIGNMENTS

RESULT	ID	035902	PRELIMINARY	PRT	993 AA.
AC	035902	01-JAN-1998 (TREMBLREL. 05, CREATED)			
DT	01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)				
DT	01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)				
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)				
DE	DESMOGLEIN 3 (FRAGMENT).				
GN	DSG3.				
OS	MUS MUSCULUS (MOUSE).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;				
SC	SCITOGNATHI; MURIDAE; MURINAE; MUS.				
RP	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-BALB/C;				
RA	ISHIKAWA H., LI K., UETTO J.;				
RL	SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.				
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).				
DR	EMBL: U86016; G2290200; -				
DR	PROSITE: P500232; CADHERIN; 2.				
DR	PFAM: PF00028; cadherin; 4.				
KW	CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.				
FT	NON_TER 993				
SO	SEQUENCE 993 AA; 107888 MW; 881794BD CRC32;				

Query Match 91.1%; Score 92; DB 11; Length 993;
Best Local Similarity 100.0%; Pred. No. 8.44e-09;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	81 OKITRISGVGID 93
QY	3 OKITRISGVGID 15
RESULT	2 PRELIMINARY; PRT; 3380 AA.
ID	Q24292
AC	024292;
DT	01-NOV-1996 (TREMBLREL. 01, CREATED)
DT	01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE	ADHERIN.
GN	DACHSUS.
OS	DROSOPHILA MELANOGASTER (FRUIT FLY).
OC	EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
SC	PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;

```

OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95324812.
RA CLARK H.F., BRENNERUP D., SCHNEITZ K., BIBER A., GOODMAN C., NOLL M.;
RT "Drosophila encodes a member of the cadherin superfamily that controls
RT imaginal disc morphogenesis in Drosophila."
RL GENES DEV. 9:1530-1542(1995).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; L08811; G685170; -.
DR FLYBASE; FBgn0000497; ds.
DR PROSITE; PS00232; CADHERIN; 18.
DR PFAM; PF00028; cadherin; 26.
KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
SQ SEQUENCE 3380 AA; 366356 MW; 7D83FAC3 CRC32;

Query Match 59.4%; Score 60; DB 5; Length 3380;
Best Local Similarity 58.3%; Pred. No. 3,90e-01;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 1017 RITYRIYVAGVD 1028
OY 4 KITRIISGVGID 15

RESULT 3 PRELIMINARY; PRT; 2269 AA.
ID 077360.
AC 077360.
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE MA3P4.25 PROTEIN.
GN MA3P4.25.
OS PLASMODIUM FALCIPARUM.
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-3D7;
RA HAMLIN N., LAMSON D., BARRELL B.;
RL SUBMITTED (OCT-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AL008970; E1332566; -.
SQ SEQUENCE 2269 AA; 267233 MW; 11E1C8E0 CRC32;

Query Match 58.4%; Score 59; DB 5; Length 2269;
Best Local Similarity 88.9%; Pred. No. 6,36e-01;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 573 OXIGYRISG 581
OY 3 OXITIRISG 11

RESULT 4 PRELIMINARY; PRT; 331 AA.
ID 066968.
AC 066968.
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE PHOSPHORIBOSYLFORMYLGLYCINAMIDINE CYCLO-LIGASE.
GN PUBM.
OS AQUIFEX AEOLICUS.
OC BACTERIA; AQUIFICALES; AQUIFICACEAE; AQUIFEX.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-VF5;
RX MEDLINE; 98196666.
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., AJAYI M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL NATURE 392:353-358(1998).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., AJAYI M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AE000706; G2983329; -.
KW LIGASE.
SQ SEQUENCE 331 AA; 36768 MW; 4C3D13B0 CRC32;

Query Match 57.4%; Score 58; DB 2; Length 331;
Best Local Similarity 58.3%; Pred. No. 1,03e+00;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1 MYTYSAGVD 12
OY 4 KITRIISGVGID 15

RESULT 5 PRELIMINARY; PRT; 441 AA.
ID 022282.
AC 022282.
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DE T517.1 OR F17A14.8.
GN T517.1 OR F17A14.8.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTES; EUDICOTYLEDONS; ROSIDAE;
OC EUPHYLLIPHYTES; SPERMATOPHYTES; MAGNOLIOPHYTES; EUDICOTYLEDONS; ROSIDAE;
OC CAPRIFALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AC003000; G2642153; -.
DR EMBL; AC003674; G2795810; -.
SQ SEQUENCE 441 AA; 47873 MW; F462C421 CRC32;

Query Match 56.4%; Score 57; DB 10; Length 441;
Best Local Similarity 40.0%; Pred. No. 1.66e+00;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 366 SQKLLYRVNGAGSE 380
OY 1 AKKITRIISGVGID 15

RESULT 6 PRELIMINARY; PRT; 1363 AA.
ID 066291.
AC 066291.
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DE SURFACE PROTEIN PRECURSOR.
GN HDAN CORONAVIRUS.
OS VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; NIDOVIRALES;
OC CORONAVIRIDAE; CORONAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-HCV-OC43;
RA KUENKELE F., HERRLER G.;
RL ARCH. VIROL. 141:1123-1131(1996).
RN [2]

```

DR EMBL: 232769; G475894; .
KM SIGNAL. 1 17 POTENTIAL.
FT CHAIN 18 1363 SURFACE PROTEIN.
SQ SEQUENCE 1363 AA; 150835 MM; A280BD88 CRC32;
Query Match 56.4%; Score 57; DB 14; Length 1363;
Best Local Similarity 45.5%; Pred. No. 1.66e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 995 NVOYRINGIGV 1005
QY 4 KITYRISGVGI 14
RESULT 7
ID 066290 PRELIMINARY; PRT; 1363 AA.
AC 066290;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 03, LAST ANNOTATION UPDATE)
DE SURFACE PROTEIN PRECURSOR.
OS HUMAN CORONAVIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; NIDOVIRALES;
OC CORONAVIRIDAE; CORONAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OC43;
RA KUENKELE F., HERRLER G.;
RL ARCH. VIROL. 1411:1123-1131(1996).
DR EMBL: 232768; G475892; .
KM SIGNAL. 1 17 POTENTIAL.
FT CHAIN 18 1363 SURFACE PROTEIN.
SQ SEQUENCE 1363 AA; 150701 MM; 4510BD9B CRC32;
Query Match 56.4%; Score 57; DB 14; Length 1363;
Best Local Similarity 45.5%; Pred. No. 1.66e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 995 NVOYRINGIGV 1005
QY 4 KITYRISGVGI 14
RESULT 8
ID 051462 PRELIMINARY; PRT; 419 AA.
AC 051462;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 48.7 KD PROTEIN.
GN BB0509
OS BORRELIA BURGDORFERI (LYME DISEASE SPIROCHETE).
OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; BORRELIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 35210 / B31;
RX MEDLINE: 98065943
RA FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
RA LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
RA PETERSON J., KRALAVAGE A.R., QUACKENBUSH J., SALBERG S., HANSON M.,
RA VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
RA UTTERBACK T., WATHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
RA SMITH H.O., VENTER J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
burgdorferi."
RT NATURE 390:580-586(1997).
DR EMBL: A8001153; G2688429; .
KM TIGR: BB0509; .
HYPOTHETICAL PROTEIN.

SQ SEQUENCE 419 AA; 48741 MM; 7389D9CF CRC32;
Query Match 55.4%; Score 56; DB 2; Length 419;
Best Local Similarity 38.5%; Pred. No. 2.67e+00;
Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;
Db 278 RNIAVRININVD 290
QY 3 OKITIRISGVGI 15
RESULT 9
ID 041278 PRELIMINARY; PRT; 108 AA.
AC 041278;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE STRAMBERY MILD YELLOW EDGE-ASSOCIATED POTEXVIRUS COAT PROTEIN,
DE PARTIAL REPLICASE AND ORF2, ORF3, ORF4
OS STRAMBERY MILD YELLOW EDGE-ASSOCIATED VIRUS (SMYEV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTEXVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D-74;
RA LAMPRECHT S., JELKMAN W.;
RL ACTA HORTICULTURAE 0:0-0(0).
DR EMBL: Y13938; E324688; .
DR PFAM: PF01307; Plant_vir_prot; 1.
SQ SEQUENCE 108 AA; 11613 MM; 1D982A03 CRC32;
Query Match 54.5%; Score 55; DB 14; Length 108;
Best Local Similarity 38.3%; Pred. No. 4.25e+00;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 9 HSTYRIYAVGL 20
QY 3 OKITIRISGVGI 14
RESULT 10
ID 020481 PRELIMINARY; PRT; 287 AA.
AC 020481;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE F46G10.3 PROTEIN.
GN F46G10.3
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA; RHABDITIDAE;
OC RHABDITIDAE; RHABDITIDAE; RHABDITIDAE; RHABDITIDAE; RHABDITIDAE;
RN [1]
RP SEQUENCE FROM N.A.
RA COLES L.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEKES M.,
RA BONFIELD J., BORTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RITKEN L., ROOPRA A., SAUNDERS D., SHANKEN R.,
RA SWALDON N., SMITH A., SONNHAMER E., STADEN K., SUTSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN R., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPOAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans".
RT NATURE 368:32-38(1994).
DR EMBL: Z50177; E1346901; .
SQ SEQUENCE 287 AA; 32464 MM; F4C7797C CRC32;

Query Match 53.5%; Score 54; DB 5; Length 287;
Best Local Similarity 66.7%; Pred. No. 6,74e+00;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 272 ATMKLDYRISDV 283
1 ATOKITRISGV 12

RESULT 11
ID 059326 PRELIMINARY; PRT; 482 AA.
AC 059326;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CHITINASE PRECURSOR (EC 3.2.1.14) (CHITODEXTRINASE)
DE (1,4-BETA-POLY-N-ACETYLGLUCOSAMINIDASE) (POLY-BETA-GLUCOSAMINIDASE).
GN CHIT.
OS CLOSTRIDIUM THERMOCELLOM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 1237;
RL FUCHS K.P., SCHWARZ W.H., STAUDENBAUER W.L.;
RA SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF
N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
DR EMBL: 268924; E220269;
DR PROSITE: PS00448; CLOS_CELULOSOME_RPT; 2.
DR PROSITE: PS01095; CHITINASE_18; 1.
DR PEFAM: PF00704; glycosyl_hydro; 2.
DR PEFAM: PF00704; glycosyl_hydro; 2.
DR SIGNAL: HYDROLASE; GLYCOSIDASE.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 482
SQ SEQUENCE 482 AA; 55028 MW; F75DCAD0 CRC32;

Query Match 53.5%; Score 54; DB 2; Length 482;
Best Local Similarity 70.0%; Pred. No. 6,74e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DB 152 TYRFDGVID 161
6 TYRISGV 15

RESULT 12
ID 015855 PRELIMINARY; PRT; 878 AA.
AC 015855; Q16194; Q13799;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE OVOMORLIN PRECURSOR (E-CADHERIN) (ARC-1/OVOMORLIN).
GN UVO.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LIVER;
RX MEDLINE: 94242050.
RA RIMM D.L., MORROW J.S.;
RA "Molecular cloning of human E-cadherin suggests a novel subdivision
of the cadherin superfamily.";
RL BIOCHEM. BIOPHYS. RES. COMMON. 200:1754-1761(1994).
RN [2]
RN SEQUENCE OF 333-472 FROM N.A.
RX MEDLINE: 94306394.
RA BECKER K.F., ATKINSON M.J., REICH U., BECKER I., NEKARDA H.,
RA SIEMER J.R., HOLTER H.;
RA "E-cadherin gene mutations provide clues to diffuse type gastric
carcinomas.";

Query Match 54.3845-3852(1994).
Best Local Similarity 100.0%; Pred. No. 6,74e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 515 OKITRYI 521
3 OKITRYI 9

RESULT 13
ID 019482 PRELIMINARY; PRT; 2610 AA.
AC 019482; Q23218;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE F15B9.7 PROTEIN.
GN F15B9.7.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA BAYNES C.;
RN [2]
RN SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RN SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSKOGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAYTON M., DEAR S., DU Z., DUBREIN R., FAVELLO A., FULTON L.,
RA GARNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMAILDON N., SMITH A., SONNEMMER E., STADEN R., SULLSTON J.,
RA THIERREY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL NATURE 368:32-38(1994).
RN [3]
RN SEQUENCE FROM N.A.
RP PERCY C.;
RA SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: 278018; E1350097;
DR EMBL: 278013; E1350097; JOINED.
DR EMBL: 278013; E1345595;
DR EMBL: 278018; E1345595; JOINED.
DR PROSITE: PS00332; CADHERIN; 6.
KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
SQ SEQUENCE 2610 AA; 280075 MW; 1D187623 CRC32;

Query Match 53.5%; Score 54; DB 5; Length 2610;
Best Local Similarity 100.0%; Pred. No. 6.74e+00;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 612 OKITTRI 618
OY 3 OKITTRI 9

Db 69 AOKETCYCLPNIGID 82
OY 2 TOKITTRISGVGID 15

Search completed: Fri Jun 11 17:11:56 1999
Job time : 130 secs.

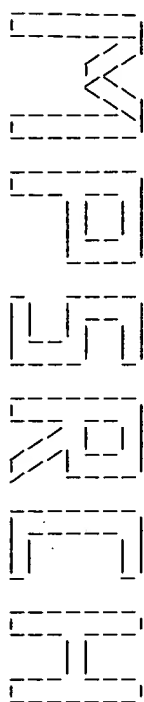
RESULT 14
ID 077862 PRELIMINARY; PRT; 82 AA.
AC 077862;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
OS MHC CLASS II B LOCUS 4 (FRAGMENT).
OC OREOCHROMIS NILOTICUS (NILE TILAPIA) (TILAPIA NILOTICA).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA; PERCIFORMES;
OC LABROIDEI; CICHLIDAE; TILAPIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 9831513.
RA MALAGA-TRILLO E., MCANDREW B., VINCEK V., ZALESKA-RUTCZYNSKA Z.,
RA SUELTSMANN H., FIGUEROA F., KLEIN J.;
RT "Linkage Relationships and Haplotype Polymorphism among Cichlid MHC
RT class II B genes";
RL GENETICS 146:1527-1547(1998).
DR EMBL; AF049971; G3282889; -.
KM MHC.
FT NON_TER 1 1
FT NON_TER 82 82
SQ SEQUENCE 82 AA; 9607 MM; D1718622 CRC32;

Query Match 52.5%; Score 53; DB 7; Length 82;
Best Local Similarity 50.0%; Pred. No. 1.06e+01;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 69 AOKETCYCLPNIGID 82
OY 2 TOKITTRISGVGID 15

RESULT 15
ID 031536 PRELIMINARY; PRT; 82 AA.
AC 031536;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MHC CLASS II BETA CHAIN (FRAGMENT).
GN 2C.
OS PSEUDOTROPHUS ZEBRA.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA; PERCIFORMES;
OC LABROIDEI; CICHLIDAE; PSEUDOTROPHUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94018589.
RA ONO H., O'HUIGIN C., TICHY H., KLEIN J.;
RT "Major histocompatibility-complex variation in two species of cichlid
RT fishes from Lake Malawi";
RL MOL. BIOL. EVOL. 10:1060-1072(1993).
DR EMBL; L17460; G309928; -.
DR PFAM; PF00969; MHC_II_beta; 1.
KM MHC.
FT NON_TER 1 1
FT NON_TER 82 82
SQ SEQUENCE 82 AA; 9754 MM; 90E5E756 CRC32;

Query Match 52.5%; Score 53; DB 7; Length 82;
Best Local Similarity 50.0%; Pred. No. 1.06e+01;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;



(TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_dp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 11 17:20:38 1999; MasPar time 4.88 Seconds

Tabular output not generated. 65.330 Million cell updates/sec

Title: >US-08-991-628-2

Description: (1-15) from US08991628.pep

Sequence: 1 FGIFVVDKNTGNDINI 15

Scoring table: PAM 150

Gap 15

Searched: 170751 seqs; 21266608 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 18.438; Variance 59.855; scale 0.308

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	105	100.0	15	20	M04842	Self epitope of desmo
2	105	100.0	30	16	R93960	Peptide combining wit
3	105	100.0	614	19	W07908	Pemphigus vulgaris an
4	105	100.0	999	6	R30742	Human pemphigus vulga
5	86	81.9	19	16	R93961	Peptide combining wit
6	86	81.9	21	16	R93962	Peptide combining wit
7	84	80.0	778	21	M15489	Pemphigus foliaceus a
8	82	78.1	560	24	M13009	Segment of desmosomal
9	57	54.3	793	9	R49732	Sequence encoded by h
10	57	54.3	796	38	M85598	Cadherin-11.
11	57	54.3	796	9	R49731	Sequence encoded by h
12	57	54.3	796	9	R49730	Sequence encoded by m
13	57	54.3	796	21	M13134	Full length human cad
14	57	54.3	796	23	W25636	Human cadherin-11.
15	57	54.3	797	17	R68666	Human protocadherin p
16	56	53.3	433	30	W41732	Arabidopsis chloropla

17	56	53.3	511	13	R69506	Aspergillus sp. recom	6.44e+01
18	56	53.3	515	13	R69504	Aspergillus sp. recom	6.44e+01
19	56	53.3	515	13	R69505	Aspergillus sp. recom	6.44e+01
20	56	53.3	713	23	W25638	Human cadherin-13.	6.44e+01
21	56	53.3	713	21	M13136	Full length human cad	6.44e+01
22	55	52.4	17	16	R93963	Peptide combining wit	8.10e+01
23	55	52.4	20	29	W40250	Human wild-type E-cad	8.10e+01
24	55	52.4	499	12	R60609	Tobamovirus replicati	8.10e+01
25	55	52.4	878	15	R85487	Human E-cadherin prec	8.10e+01
26	55	52.4	878	10	R55060	Sequence of human liv	8.10e+01
27	54	51.4	246	37	W76815	N. gonorrhoeae piliC2	1.02e+02
28	54	51.4	246	5	R26789	Sequence encoded by t	1.02e+02
29	54	51.4	557	29	W43448	Tobacco lactase clone	1.02e+02
30	53	50.5	199	7	R34398	Helicobacter pylori u	1.28e+02
31	53	50.5	794	21	M13135	Putative human cadher	1.28e+02
32	53	50.5	794	23	W25637	Human cadherin-12.	1.28e+02
33	53	50.5	2647	25	M19349	Human filamin.	1.28e+02
34	52	49.5	115	34	W62680	Streptococcus pneumon	1.60e+02
35	52	49.5	507	26	W29772	Calaszeria fungus MF-	1.60e+02
36	52	49.5	535	11	R56549	Cold acclimatization	1.60e+02
37	52	49.5	1026	17	R87146	Protocadherin clone 4	1.60e+02
38	52	49.5	1026	11	R58906	Human protocadherin-4	1.60e+02
39	52	49.5	1203	17	R87152	Alternatively spliced	1.60e+02
40	52	49.5	1203	11	R58911	Product of alternativ	1.60e+02
41	51	48.6	124	31	W27945	Staphylococcus aureus	2.00e+02
42	51	48.6	409	29	W43449	Tobacco lactase clone	2.00e+02
43	51	48.6	712	4	R27823	Sequence encoded by T	2.00e+02
44	51	48.6	717	4	R27824	Sequence encoded by T	2.00e+02
45	50	47.6	251	22	W20277	H. pylori surface or	2.50e+02

ALIGNMENTS

RESULT 1
ID W04842: standard; peptide; 15 AA.

AC 18-FEB-1997 (first entry)
DE Self epitope of desmoglein 3, implicated in autoimmune disease.
KW Tolerisation; self-epitope; antigen; autoimmune disease;
KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
KW pemphigus vulgaris; desmoglein; multiple sclerosis;
KW herpes simplex virus; adenovirus; phosphomannomutase;
KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
KW Influenza; haemagglutinin; reovirus; sigma protein.
OS Homo sapiens.
PN W09627387-A1.
PD 12-SEP-1996
PF 07-MAR-1996; U03182.
PR 07-MAR-1995; US-400796.
PA (HARD) HARVARD COLLEGE.
PI Strominger JL, Muecherfennig KW;
DR WPI: 96-425218/42.
PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
PT antigens - useful in disease treatment, and method for
PT identification of other self and non-self antigens implicated in
PT auto-immune disease
PS Claim 1; Page 38; 58pp; English.
CC Pharmaceutical preparations for tolerisation to antigens comprise
CC either an isolated human non-collagen or non-mwsln basic protein
CC (MBP) polypeptide which is capable of tolerising an individual to an
CC autoantigen; or an isolated human pathogen polypeptide capable of
CC tolerising an individual to that polypeptide. In both cases, the
CC polypeptide (whether self or non-self) includes an amino acid
CC sequence corresponding to a sequence motif for a MHC class II
CC protein, such as HLA-DR, which is associated with a human autoimmune
CC disease and which binds to the polypeptide to activate autoreactive
CC T-cells in individuals with the autoimmune disease. This peptide is
CC derived from the human desmoglein 3 protein (amino acids 97-111)
CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
CC derived from the human desmoglein protein are described in W04841-47.
SQ Sequence 15 AA;

Query Match 100.0%; Score 105; DB 20; Length 15;

Best Local Similarity 100.0%; Pred. No. 2,95e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 fgifvdkntgdlnl 15
| | | | | | | | | | |
Oy 1 FGIFVVKNTGDINI 15

RESULT 2
ID R93960 standard; peptide; 30 AA.
AC R93960;

DE 05-JUL-1996 (first entry)
DE Peptide combining with anti-interferidermal cellular antibody;
KW anti-interferidermal cellular antibody; autoantibody; adsorbent.

OS Synthetic.
PN J07309893-A.
PD 28-NOV-1995.
PE 18-MAY-1994; 129556.
PR 18-MAY-1994; JP-129556.
PA (KURS) KURARAY CO LTD.
WP: 96-045392/05.

DR Anti-interferidermal cellular antibody-combining peptide - which can
PT be immobilised on column to form adsorbent useful for treating
PS diseases related to the antibody

PS Claim 1: Page 2; 7pp; Japanese.
CC A new peptide is disclosed which contains at least 5 contiguous amino
CC acids from the sequence Val Gly Ile Asp Gln Pro Phe Gly Ile Phe
CC Val Val Asp Lys Asn Thr Gly Asp Ile Asn Ile Thr Ala Ile Val Asp Arg
CC Glu Glu (the present sequence), the peptide not containing more than
CC 50 residues. The peptide combines with anti-interferidermal cellular
CC antibody. It can be immobilised on a carrier to prepare an adsorbent
CC useful for the treatment of diseases related to anti-interferidermal
CC cellular antibody.
SQ Sequence 30 AA;

Query Match 100.0%; Score 105; DB 16; Length 30;
Best Local Similarity 100.0%; Pred. No. 2,95e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 8 fgifvdkntgdlnl 22
| | | | | | | | | | |
Oy 1 FGIFVVKNTGDINI 15

RESULT 3
ID W07908 standard; protein; 614 AA.
AC W07908;

DE 29-JAN-1997 (first entry)
DE Pemphigus vulgaris antigen protein extracellular region.
KW Autoantibody; immunoglobulin G; IgG1; fusion protein; diagnosis;
KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;
KW dermatology.

OS Homo sapiens.
PN J08188540-A.
PD 23-JUL-1996.
PE 30-JUN-1995; 165632.
PR 30-JUN-1994; JP-173291.
PA (NISH) NISHUKAWA T.
WP: 96-388562/39.

DR Fused protein recognised by pemphigus vulgaris autoantibody -
PT useful to treat and diagnose pemphigus vulgaris

PS Claim 1: Page 7-9; 9pp; Japanese.
CC W07908 represents the human pemphigus vulgaris (PV) antigen
CC extracellular region. The PV antigen is produced in patients with
CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare
CC relapsing disease causing suprabasal, intra-epidermal bullae
CC (vesicles) of the skin and mucous membranes, which is fatal if
CC untreated. The PV antigen was fused to a human IgG1 hinge region
CC and the resulting fusion protein is useful to treat or diagnose
CC pemphigus vulgaris.
SQ Sequence 614 AA;

Query Match 100.0%; Score 105; DB 19; Length 614;

Best Local Similarity 100.0%; Pred. No. 2,95e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 96 fgifvdkntgdlnl 110
| | | | | | | | | | |
Oy 1 FGIFVVKNTGDINI 15

RESULT 4
ID R30742 standard; Protein; 999 AA.
AC R30742;

DE 14-JUN-1993 (first entry)
DE Human pemphigus vulgaris 130kD antigen;
KW Pemphigus vulgaris; skin disease; autoantibodies;
KW keratinocyte cell surface antigen; glycoprotein; cell adhesion.

OS Homo sapiens.
PN US7798918-A.
PD 15-DEC-1992.
PE 27-NOV-1991; 798918.
PR (USSH) US DEPT HEALTH & HUMAN SERVICE.
PA Amagal M, Klaus-Kovtun V, Stanley JR;
DR WP: 93-067436/08.

DR N-PSDB: 035992.
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
PT diagnostic and therapeutic uses

PS Disclosure; Fig 7; 50pp; English.
CC This sequence is the pemphigus vulgaris 130kD antigen. The protein
CC and its encoding DNA may be used in the diagnosis and treatment of
CC pemphigus vulgaris. It is thought that the antigen may be a cell
CC adhesion molecule.
SQ Sequence 999 AA;

Query Match 100.0%; Score 105; DB 6; Length 999;
Best Local Similarity 100.0%; Pred. No. 2,95e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 97 fgifvdkntgdlnl 111
| | | | | | | | | | |
Oy 1 FGIFVVKNTGDINI 15

RESULT 5
ID R93961 standard; peptide; 19 AA.
AC R93961;

DE 05-JUL-1996 (first entry)
DE Peptide combining with anti-interferidermal cellular antibody.
KW anti-interferidermal cellular antibody; autoantibody; adsorbent.
OS Synthetic.
PN J07309893-A.
PD 28-NOV-1995.
PE 18-MAY-1994; 129556.
PR 18-MAY-1994; JP-129556.
PA (KURS) KURARAY CO LTD.
WP: 96-045392/05.

DR Anti-interferidermal cellular antibody-combining peptide - which can
PT be immobilised on column to form adsorbent useful for treating
PS diseases related to the antibody

PS Example 1: Page 4; 7pp; Japanese.
CC New peptides are disclosed which contain at least 5 contiguous amino
CC acids from the sequence Val Gly Ile Asp Gln Pro Phe Gly Ile Phe
CC Val Val Asp Lys Asn Thr Gly Asp Ile Asn Ile Thr Ala Ile Val Asp Arg
CC Glu Glu (see R93960), the peptide not containing more than 50 residues.
CC The peptide combines with anti-interferidermal cellular antibody. It can
CC be immobilised on a carrier to prepare an adsorbent useful for the
CC treatment of diseases related to anti-interferidermal cellular antibody.
CC The present sequence is a specific example of the new peptides.
SQ Sequence 19 AA;

Query Match 81.9%; Score 86; DB 16; Length 19;
Best Local Similarity 100.0%; Pred. No. 4,18e-02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 8 fgifvdkntgd 19
 |||||||
 QY 1 fgifvdkntgd 12

RESULT 6
 ID R93962 standard; peptide: 21 AA.

AC R93962: (first entry)
 DT 05-JUL-1996
 DE Peptide combining with anti-interepidermal cellular antibody.
 KW anti-interepidermal cellular antibody; autoantibody; adsorbent.
 OS Synthetic.

PN J07309893-A.
 PD 28-NOV-1995.
 PF 18-MAY-1994; 129556.
 PR 18-MAY-1994; JP-129556.
 PA (KURS) KURARAY CO LTD.
 DR WPI; 96-045392/05.

PT Anti-interepidermal cellular antibody-combining peptide - which can be immobilised on column to form adsorbent useful for treating diseases related to the antibody.
 PS Example 2; Page 5; 7pp; Japanese.

CC New peptides are disclosed which contain at least 5 contiguous amino acids from the sequence Val Gly Ile Asp Gln Pro Phe Gly Ile Phe Val Val Asp Lys Asn Thr Gly Asp Ile Asn Thr Ala Ile Val Asp Arg Glu Glu (see R93960), the peptide not containing more than 50 residues. The peptide combines with anti-interepidermal cellular antibody. It can be immobilised on a carrier to prepare an adsorbent useful for the treatment of diseases related to anti-interepidermal cellular antibody.
 CC The present sequence is a specific example of the new peptides.
 SO Sequence 21 AA;

Query Match 81.9%; Score 86; DB 16; Length 21;
 Best Local Similarity 100.0%; Pred. No. 4.18e-02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 10 fgifvdkntgd 21
 |||||||
 QY 1 fgifvdkntgd 12

RESULT 7
 ID W15489 standard; Protein: 778 AA.

AC W15489: (first entry)
 DT 17-JUN-1997
 DE Pemphigus foliaceus antigen-IgG constant region fusion protein.
 KW Pemphigus foliaceus; autoantibody; constant region; IgG;
 KW extracellular region; antigen; hinge portion; skin;
 KW dermatitis herpetiformis; fusion protein; detection; ss.
 OS Chimeric - Homo sapiens.

FT Key Location/Qualifiers
 FT domain 1..545
 FT /note= "Pemphigus foliaceus antigen protein"
 PN J09077800-A.
 PD 25-MAR-1997.
 PF 12-SEP-1995; 260899.
 PR 12-SEP-1995; JP-260899.
 PA (NISH/) NISHIKAWA T.
 DR WPI; 97-241758/22.
 DR P-PDB; T66428.

PT Pemphigus foliaceus antigen-IgG constant region fusion protein - linked through the hinge region used to treat pemphigus foliaceus
 PS Claim 1; Page 10-12; 17pp; Japanese.
 CC This sequence represents a fused protein recognised by pemphigus foliaceus patient autoantibody which comprises the constant region of IgG linked to the extracellular region of pemphigus foliaceus antigen protein through the hinge portion. Pemphigus foliaceus is a chronic, generalised, vesicular and scaling skin eruption similar to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion protein is useful to treat pemphigus foliaceus. The antigen is especially administered through an adsorbent upon which the fusion protein is immobilised via a carrier. The fusion protein is also useful for detecting pemphigus foliaceus antibodies which is useful

CC in immunodiagnosis. The fusion protein has little or no side effects.
 SQ Sequence 778 AA;

Query Match 80.0%; Score 84; DB 21; Length 778;
 Best Local Similarity 60.0%; Pred. No. 6.96e-02;
 Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 97 ygfivnktgetnl 111
 :||||:|||||
 QY 1 fgifvdkntgd 15

RESULT 8
 ID W13009 standard; Protein: 560 AA.

AC W13009:
 DT 21-NOV-1997 (first entry)
 DE Segment of desmosomal cadherin, desmoglein Dsg2.
 KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
 KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;
 KW micrometastasis; separation; enrichment; targeted delivery;
 KW metastatic.

OS Homo sapiens.
 PN DE19531033-A1.
 PD 27-FEB-1997.
 PR 23-AUG-1995; 031033.
 PF 23-AUG-1995; DE-031033.
 PA (PROG-) PROGEN BIOTECHNIK GMBH.
 PI Franke WW, Schaefer S;
 DR WPI; 97-146518/14.

PT Antibody reactive with part of desmosomal cadherin - exposed on surface of epithelial or carcinoma cells, not bound to desmosomes, useful for diagnosis and treatment of carcinoma micrometastases
 PS Claim 7; Page 5; 8pp; German.
 CC The present sequence is a segment of the desmosomal cadherin (DC), desmoglein Dsg2, which is exposed on the surface of epithelial or carcinoma cells and not bound to desmosomes. An antibody (Ab) directed against epitopes of the present sequence can be used to diagnose, i.e. to detect carcinoma cells, especially micrometastases, not bound to desmosomes, to separate, enrich or detect living or fixed carcinoma cells by cell sorting methods and as a therapeutic to deliver agents, e.g. other Ab or toxins, to target cells. The Ab provides rapid and reliable detection of metastatic carcinoma, and detects parts of DC that are not accessible in desmosome bound cells, as in normal tissue or carcinomas.
 CC Sequence 560 AA;

Query Match 78.1%; Score 82; DB 24; Length 560;
 Best Local Similarity 60.0%; Pred. No. 1.16e-01;
 Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

DB 50 fgifvdkntgetnl 64
 |||||:|||||
 QY 1 fgifvdkntgd 15

RESULT 9
 ID R49732 standard; Protein: 693 AA.

AC R49732:
 DT 14-SEP-1994 (first entry)
 DE Sequence encoded by human OSF-4-2 cDNA.
 KW OSF-4-2; cadherin; growth factor; osteogenesis; osteoblast; therapy;
 KW diagnosis.

OS Homo sapiens.
 PN EP-585801-A.
 PD 09-MAR-1994.
 PR 25-AUG-1993; 113602.
 PF 28-AUG-1992; JP-230028.
 PA (FARR) HOECHST JAPAN LTD.
 PI Amann E, Kawai S, Okazaki M, Takeshita S, Tsujimura A;
 DR WPI; 94-076152/10.
 DR N-PSDB; Q44393.
 PT New bone related, cadherin-like OSF-4 proteins - for treatment

PT and diagnosis of bone metabolic disease, and nucleic acid
 encoding them
 PS Claim 1: Page 23-27: 34pp; English.
 CC cDNA libraries were constructed from the mouse osteoblastic cell
 line MC3T3-E1 and from mouse liver tissue, amplified by PCR, and
 then as much common DNA as possible removed by hybridisation between
 the 2 libraries. Residual E1-specific DNA was amplified, inserted
 into lambda gt10 and screened by plaque hybridisation. A minbank of
 273 E-specific clones was recovered, their inserts amplified and
 used to screen total RNA from both cell types. One clone specific
 for E1 was identified and sequenced. The insert from this clone was
 used to screen cDNA prep. from E1 RNA and the longest posn. insert
 cloned in pGEM 112f (+) to give pKOR164. This insert was sequenced;
 it encoded the 796 AA mouse precursor protein (Q44391/R49730). The
 insert was also used to screen a cDNA bank prep. from human
 osteosarcoma to identify 2 clones encoding the 2 human precursor
 proteins - OSF-4-1 and OSF-4-2 (Q44392/R49731 and Q44393/R49732
 respectively).
 SQ Sequence 693 AA;

Query Match 54.3%; Score 57; DB 9; Length 796;
 Best Local Similarity 50.0%; Pred. No. 5.11e+01;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 100 lfvdskgnh 111
 |||:|:|:|:
 OY 3 IFVVDKNTGGIN 14

RESULT 10
 ID W85598 standard; protein: 796 AA.
 AC W85598;
 DT 02-MAR-1999 (first entry)
 DE Cadherin-11.
 KW Cadherin; morphogenesis; calcium-dependent cell adhesion; pregnancy;
 endometrium; trophoblast; blastocyst; infertility.
 OS Homo sapiens.
 PN W09849560-A1.
 PD 05-NOV-1998.
 PF 24-APR-1998: CA0397.
 PR 25-APR-1997: CA-203718.
 RA (UYBR-) UNIV BRITISH COLUMBIA.
 PI Macalman CD, Stephenson MD;
 DR WPI: 99-024071/02.
 DR N-PSDB: V83124.
 PT Assessing likelihood of successful pregnancy by measuring levels of
 cadherin-11 in endometrium - also diagnosis of infertility from low
 cadherin levels and increasing cadherin levels by administering
 progesterin or cadherin-encoding DNA
 PS Disclosure: Page 54-57: 63pp; English.
 CC The likelihood of establishing and maintaining a pregnancy, blastocyst
 implantation or endometrial receptivity are determined by measuring the
 level of cadherin-11 mRNA or protein in endometrial cells. A level below
 a standard value indicates inability to establish or maintain pregnancy.
 CC Women who are identified as having low level expression of
 cadherin-11 in endometrial cells can then be treated with a genetic
 construct comprising the cadherin-11 cDNA. The expression of
 cadherin-11 from the construct increases fertility and lessens the
 likelihood of miscarriage.
 CC Cadherin-11 expression is a better predictor of endometrial response
 and receptiveness than conventional analysis of endometrial cell
 morphology.
 SQ Sequence 796 AA;

Query Match 54.3%; Score 57; DB 38; Length 796;
 Best Local Similarity 50.0%; Pred. No. 5.11e+01;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 100 lfvdskgnh 111
 |||:|:|:|:
 OY 3 IFVVDKNTGGIN 14

RESULT 11
 ID R49731 standard; Protein: 796 AA.
 AC R49731;
 DT 14-SEP-1994 (first entry)
 DE Sequence encoded by human OSF-4-1 cDNA.
 KW OSF-4-1; cadherin; growth factor; osteogenesis; osteoblast; therapy;
 diagnosis.
 OS Homo sapiens.
 PN EP-585801-A.
 PD 09-MAR-1994.
 PF 25-AUG-1993; 113602.
 PR 28-AUG-1992; JP-230028.
 RA (FARH) HOECHST JAPAN LTD.
 PI Amann E, Kawai S, Okazaki M, Takeshita S, Tsujimura A;
 DR WPI: 94-076152/10.
 DR N-PSDB: Q44392.
 PT New bone related, cadherin-like OSF-4 proteins - for treatment
 and diagnosis of bone metabolic disease, and nucleic acid
 encoding them
 PS Claim 1: Page 18-22: 34pp; English.
 CC cDNA libraries were constructed from the mouse osteoblastic cell
 line MC3T3-E1 and from mouse liver tissue, amplified by PCR, and
 then as much common DNA as possible removed by hybridisation between
 the 2 libraries. Residual E1-specific DNA was amplified, inserted
 into lambda gt10 and screened by plaque hybridisation. A minbank of
 273 E-specific clones was recovered, their inserts amplified and
 used to screen total RNA from both cell types. One clone specific
 for E1 was identified and sequenced. The insert from this clone was
 used to screen cDNA prep. from E1 RNA and the longest posn. insert
 cloned in pGEM 112f (+) to give pKOR164. This insert was sequenced;
 it encoded the 796 AA mouse precursor protein (Q44391/R49730). The
 insert was also used to screen a cDNA bank prep. from human
 osteosarcoma to identify 2 clones encoding the 2 human precursor
 proteins - OSF-4-1 and OSF-4-2 (Q44392/R49731 and Q44393/R49732
 respectively).
 SQ Sequence 796 AA;

Query Match 54.3%; Score 57; DB 9; Length 796;
 Best Local Similarity 50.0%; Pred. No. 5.11e+01;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 100 lfvdskgnh 111
 |||:|:|:|:
 OY 3 IFVVDKNTGGIN 14

RESULT 12
 ID R49730 standard; Protein: 796 AA.
 AC R49730;
 DT 14-SEP-1994 (first entry)
 DE Sequence encoded by murine OSF-4 cDNA.
 KW OSF-4; cadherin; growth factor; osteogenesis; osteoblast; therapy;
 diagnosis.
 OS Mus musculus.
 PN EP-585801-A.
 PD 09-MAR-1994.
 PF 25-AUG-1993; 113602.
 PR 28-AUG-1992; JP-230028.
 RA (FARH) HOECHST JAPAN LTD.
 PI Amann E, Kawai S, Okazaki M, Takeshita S, Tsujimura A;
 DR WPI: 94-076152/10.
 DR N-PSDB: Q44391.
 PT New bone related, cadherin-like OSF-4 proteins - for treatment
 and diagnosis of bone metabolic disease, and nucleic acid
 encoding them
 PS Claim 1: Page 13-17: 34pp; English.
 CC cDNA libraries were constructed from the mouse osteoblastic cell
 line MC3T3-E1 and from mouse liver tissue, amplified by PCR, and
 then as much common DNA as possible removed by hybridisation between
 the 2 libraries. Residual E1-specific DNA was amplified, inserted
 into lambda gt10 and screened by plaque hybridisation. A minbank of
 273 E-specific clones was recovered, their inserts amplified and
 used to screen total RNA from both cell types. One clone specific

CC for E1 was identified and sequenced. The insert from this clone was
 CC used to screen cDNA prepd. from E1 RNA and the longest posn. insert
 CC cloned in pGEM 112f (+) to give PKO164. This insert was sequenced;
 CC it encoded the 796 AA mouse precursor protein (Q44391/R49730). The
 CC insert was also used to screen a cDNA bank prepd. from human
 CC osteosarcoma to identify 2 clones encoding the 2 human precursor
 CC proteins - OSF-4-1 and OSF-4-2 (Q44392/R49731 and Q44393/R49732
 CC respectively).
 SQ Sequence 796 AA;

Query Match 54.3%; Score 57; DB 9; Length 796;
 Best Local Similarity 50.0%; Pred. No. 5.11e+01;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 100 Ifvldksgn111
 |||:|:|:|:|:
 Qy 3 IFVVDKNTGDN 14

RESULT 13
 ID W13134 standard; Protein; 796 AA.
 AC W13134:
 DE 14-MAY-1997 (first entry)
 DF Full length human cadherin-11.
 KW Ca2+ dependent; cell adhesion protein; foetal; cadherin; rat;
 KW brain; human; antibody; purification; determination;
 KW tissue expression; binding antagonist; calcium ion.
 OS Homo sapiens.
 PN USS597725-A.
 PD 28-JAN-1997.
 PE 17-APR-1992; 872643.
 PR 17-APR-1992; US-872643.
 PR 19-APR-1993; US-049460.
 PR 26-JAN-1994; US-188228.
 PA (DOHE-) DOHENY EYE INST.
 PI Suzuki S;
 DR WPI; 97-108328/10.
 DR N-PSDB; T61925.
 PT Antibodies to cadherin proteins - useful as cadherin antagonists,
 PT etc.
 PS Example 2: Columns 95-100; 59pp; English.
 CC The present sequence is full length human cadherin-11, which
 CC is a Ca2+ dependent cell adhesion protein. The human cadherin cDNA
 CC was isolated from a foetal brain cDNA library, using probes based
 CC on homologous rat cadherin cDNA.
 CC Antibodies or fragments that specifically bind the human cadherin
 CC can be used to purify the cadherin, determine its tissue expression
 CC and antagonise its ligand/antiligand binding activities.
 SQ Sequence 796 AA;

Query Match 54.3%; Score 57; DB 21; Length 796;
 Best Local Similarity 50.0%; Pred. No. 5.11e+01;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 100 Ifvldksgn111
 |||:|:|:|:|:
 Qy 3 IFVVDKNTGDN 14

RESULT 14
 ID W25636 standard; Protein; 796 AA.
 AC W25636:
 DE 03-NOV-1997 (first entry)
 DF Human cadherin-11.
 KW Human; cadherin; rat; calcium-dependent cell adhesion protein;
 KW superfamily; cytoskeleton; eatenin; cancer.
 OS Homo sapiens.
 PN US5646250-A.
 PD 08-JUL-1997.
 PE 17-APR-1992; 872643.
 PR 19-APR-1993; US-049460.
 PR 17-APR-1992; US-872643.
 PR 01-NOV-1994; US-332638.

PA (DOHE-) DOHENY EYE INST.
 PI Suzuki S;
 DR WPI; 97-362997/33.
 DR N-PSDB; T85403.
 PT Human and rat cadherin polypeptide(s) - mediate cell-cell adhesion
 PS Claim 1; Column 89-94; 56pp; English.
 CC This sequence represents human cadherin-11. The invention specifically
 CC provides details of human cadherin-5, -8, -11, -12 and -13, and rat
 CC cadherin-8, -5, -11 and -13. Cadherins are calcium-dependent cell
 CC adhesion proteins. They are glycosylated integral membrane proteins
 CC that have an N-terminal extracellular domain that determines binding
 CC specificity, a hydrophobic membrane spanning region and a C-terminal
 CC cytoplasmic domain, which is highly conserved among members of the
 CC superfamily. The C-terminal domain interacts with the cytoskeleton
 CC through eatenins and other cytoskeleton-associated proteins. The
 CC novel cadherin proteins may be used in the analysis of the role of
 CC cadherins in various cancers. Sequence analysis of the cadherin
 CC proteins also allows investigation of the structure and function of
 CC cadherin. The cadherin proteins may also be isolated by using anti-cadherin
 CC antibodies. These antibodies may also be used to modulate the activity
 CC of cadherin and to determine the tissue specific distribution of cadherin
 CC proteins. Each subclass of cadherins has a unique tissue distribution
 CC pattern.
 SQ Sequence 796 AA;

Query Match 54.3%; Score 57; DB 23; Length 796;
 Best Local Similarity 50.0%; Pred. No. 5.11e+01;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 100 Ifvldksgn111
 |||:|:|:|:|:
 Qy 3 IFVVDKNTGDN 14

RESULT 15
 ID R86866 standard; Protein; 797 AA.
 AC R86866:
 DE 27-AUG-1996 (first entry)
 DF Human protocadherin pc4.
 KW Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion;
 KW catenin; therapy.
 OS Homo sapiens.
 PN W09600289-A1.
 PD 04-JAN-1996.
 PE 26-JUN-1995; U08071.
 PR 27-JUN-1994; US-268161.
 PA (DOHE-) DOHENY EYE INST.
 PI Suzuki S;
 DR WPI; 96-068873/07.
 DR N-PSDB; T03573.
 PT Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
 PT pc5 - involved in cell-cell adhesion and regulation activities
 PS Claim 16; Page 122-125; 146pp; English.
 CC R86865-R86867 represent the sequences for three protocadherins. This
 CC sequence represents the human protocadherin pc4. These sequences are
 CC related to cadherin, and possess cell adhesive ability. Cadherins are
 CC glycosylated integral membrane proteins that are involved in cell-cell
 CC adhesion. Cadherins are composed of an N-terminal extracellular domain
 CC which consists of 5 unique subdomains, a membrane spanning domain, and a
 CC C-terminal cytoplasmic domain. The cytoplasmic domain interacts with the
 CC cytoskeleton through catenins and other cytoskeleton associated proteins.
 CC The cytoplasmic domain is not present in all cadherins, but in those
 CC which possess it, it is essential for the cadherin adhesive function.
 CC The cadherins which do not possess a cytoplasmic domain appear to
 CC function via a different method from those with a cytoplasmic domain.
 CC These sequences were isolated using primers 1 and 2 (see T03575 and
 CC T03576) The proteins may have regulatory functions in the cell, as well
 CC as the cell-cell adhesive properties. Antibodies produced against these
 CC sequences are useful for modulating the binding activity of these
 CC protocadherins, and can be used therapeutically.
 SQ Sequence 797 AA;

Query Match 54.3%; Score 57; DB 17; Length 797;

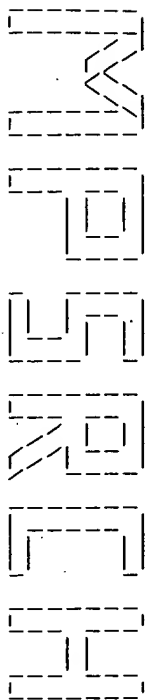
Mon Jun 14 12:18:10 1999

US-08-991-628-2.rag

Page 6

Best Local Similarity	46.7%	Pred. No. 5.1e+01	Mismatches	4	Indels	0	Gaps	0
Matches	7	Conservative						
Db	288	fpfvydelteghl	302					
					:	:	:	:
Qy	1	GFIFVDKNTGDI	15					

Search completed: Fri Jun 11 17:22:28 1999
Job time : 110 secs.


 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_p protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Jun 11 17:18:54 1999; Maspar time 4.32 Seconds
 Tabular output not generated. 139.098 Million cell updates/sec

Title: >US-08-991-628-2
 Description: (1-15) from US08991628.pep
 Perfect Score: 105
 Sequence: 1 FGIFVVDKNTGDIINI 15

Scoring table:

PAM 150
 Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database:

plrf60
 1:plrf1 2:plrf2 3:plrf3 4:plrf4

Statistics: Mean 26.793; Variance 39.116; scale 0.685

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	105	100.0	999	1	IJHUG3	desmoglein 3 precursor	1.59e-09
2	84	80.0	1043	1	IJBOG1	desmoglein 1 precursor	6.33e-05
3	84	80.0	1049	1	IJHUG1	desmoglein 1 precursor	6.33e-05
4	82	78.1	1117	2	S38673	desmoglein 2 - human	1.66e-04
5	66	62.9	761	1	IJBOD2	desmoglein 1a - bovl	2.45e-01
6	66	62.9	785	1	IJBOD2	cadherin-7 - chicken	2.45e-01
7	66	62.9	839	1	IJBOD2	desmoglein 1b precursor	2.45e-01
8	65	61.9	770	2	B48910	desmoglein 1b precursor	3.76e-01
9	65	61.9	824	2	A48910	desmoglein 1a precursor	3.76e-01
10	65	61.9	840	2	I37281	Desclb precursor - hum	3.76e-01
11	65	61.9	894	2	I37282	Desclb precursor - hum	3.76e-01
12	61	58.1	1612	2	S53969	DNA topoisomerase (AT	2.01e+00
13	61	58.1	1626	2	A39242	DNA topoisomerase (AT	2.01e+00
14	60	57.1	94	2	A65141	hypothetical 10.6 kD	3.02e+00
15	60	57.1	214	2	S76379	hypothetical protein	3.02e+00
16	60	57.1	598	2	D69292	aldehyde ferredoxin o	3.02e+00
17	60	57.1	790	2	G02678	cadherin-14 - human	3.02e+00
18	59	56.2	134	2	B64467	conserved hypothetical	4.53e+00
19	59	56.2	245	2	F64465	hypothetical protein	4.53e+00
20	59	56.2	624	2	T00044	PV72 protein - cucurb	4.53e+00
21	58	55.2	270	2	G69469	conserved hypothetical	6.75e+00
22	58	55.2	1069	2	T00040	BH-protocadherin PCDH	6.75e+00
23	58	55.2	1072	2	T00041	BH-protocadherin PCDH	6.75e+00

24	58	55.2	1200	2	T00042	BH-protocadherin PCDH	6.75e+00
25	58	55.2	3027	2	JQ1917	polyprotein - parship	6.75e+00
26	57	54.3	209	2	C64685	ribonuclease hii - He	1.00e+01
27	57	54.3	209	2	F71830	ribonuclease hii - He	1.00e+01
28	57	54.3	504	1	G71248	tLdD homolog PH0246 -	1.00e+01
29	57	54.3	796	2	I48277	cadherin-11 - mouse	1.00e+01
30	57	54.3	796	2	A38992	cadherin 11 precursor	1.00e+01
31	57	54.3	796	2	I49556	cadherin-11 - mouse	1.00e+01
32	57	54.3	796	2	A53584	OB-cadherin precursor	1.00e+01
33	57	54.3	798	2	S62791	probable lipoprotein	1.00e+01
34	57	54.3	887	1	IJCHCL	E-cadherin precursor	1.00e+01
35	57	54.3	1400	2	I38185	Protein-tyrosine kina	1.00e+01
36	56	53.3	713	2	B38992	cadherin 13 precursor	1.48e+01
37	56	53.3	756	2	S67433	hypothetical protein	1.48e+01
38	55	52.4	112	2	D64455	hypothetical protein	2.18e+01
39	55	52.4	211	2	E69391	hypothetical protein	2.18e+01
40	55	52.4	790	2	I50178	cadherin-6b - chicken	2.18e+01
41	55	52.4	884	1	IJM5CE	E-cadherin precursor	2.18e+01
42	55	52.4	1196	2	S46430	botulinum neurotoxin-	2.18e+01
43	55	52.4	1615	1	WMTM8T	180K protein - tomato	2.18e+01
44	55	52.4	1827	1	IJHUG1	sucrose alpha-glucosi	2.18e+01
45	55	52.4	1827	1	A23945	sucrose alpha-glucosi	2.18e+01

ALIGNMENTS

RESULT	ENTRY	1	ALIGNMENTS
TITLE	IJHUG3	#type complete	
ALTERNATE_NAMES	desmoglein 3 precursor - human		
ORGANISM	pemphigus vulgaris antigen		
DATE	30-Jun-1993	#sequence_revision 30-Jun-1993	#text_change
ACCESSIONS	A41088	18-Sep-1998	
REFERENCE	A41088		
authors	Amagai, M.; Kraus-Kovtun, V.; Stanley, J.R.		
Journal	Cell (1991) 67:869-877		
Title	Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a disease of cell adhesion.		
#cross-references	MU01:92069753		
#accession	A41088		
#molecule_type	MANA		
#residues	1-999	#label AMA	
#cross-references	GB:M76482; NID:9190751; PTD:9190752		
GENETICS			
gene	GDB:DSG3		
map position	18q12.1-18q12.2		
CLASSIFICATION	#superfamily cadherin; cadherin repeat homology		
KEYWORDS	calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein		
FEATURE			
1-23	#domain signal sequence #status predicted #label SIG\		
24-49	#domain propeptide #status predicted #label PRO\		
50-999	#product desmoglein homolog #status predicted #label		
50-615	#domain extracellular #status predicted #label EXT\		
51-157	#domain cadherin repeat homology #label CR1\		
160-267	#domain cadherin repeat homology #label CR2\		
270-383	#domain cadherin repeat homology #label CR3\		
390-495	#domain cadherin repeat homology #label CR4\		
496-598	#domain cadherin repeat homology #label CR5\		
616-639	#domain transmembrane #status predicted #label TM\		
640-999	#domain intracellular #status predicted #label INT\		
910-938	#domain desmoglein repeat #label DG1\		
937-965	#domain desmoglein repeat #label DG2\		
110,180,545	#binding site carbohydrate (asn) (covalent) #status predicted		
SUMMARY	#length 999 #molecular_weight 107502 #checksum 8311		
Query Match	100.0%; Score 105; DB 1; Length 999;		
Best Local Similarity	100.0%; Pred. No. 1.59e-09;		
Matches	15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		

Db 97 FGIFVVDKNTGDIINI 111
 |||||
 1 FGIFVVDKNTGDIINI 15

RESULT 2
 ENTRY IJBOGI #type complete
 TITLE desmoglein 1 precursor - bovine
 ALTERNATE_NAMES desmoglein BDCM
 ORGANISM Bos primigenius taurus #common.name cattle
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
 05-Sep-1997

ACCESSIONS S14603; A38772; A37785; S38721; A48173; S24412
 REFERENCE S14603
 #authors Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
 #submission Submitted to the EMBL Data Library, March 1991
 #description Complete sequence of the desmoglein precursor and evidence for the existence of different desmoglein genes expressed in cell type-specific patterns.

#accession S14603
 #molecule_type mRNA
 #residues 1-1043 ##label KOC
 #cross-references EMBL:X58466; NID:g306; PID:g307

REFERENCE A3872
 #authors Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.
 #journal Eur. J. Cell Biol. (1991) 55:200-208
 #title Complete amino acid sequence of the epidermal desmoglein precursor polypeptide and identification of a second type of desmoglein gene.

#cross-references M01D:92037656
 #accession A38872
 #molecule_type mRNA
 #residues 1-87:968-1043 ##label KO2
 #cross-references GB:S64268; GB:S64270

REFERENCE A37785
 #authors Goodwin, L.; Hill, J.E.; Raynor, K.; Raszi, L.; Manabe, M.; Cowin, P.
 #journal Blochem. Biophys. Res. Commun. (1990) 173:1224-1230
 #title Desmoglein shows extensive homology to the cadherin family of cell adhesion molecules.

#cross-references M01D:91097553
 #accession A37785
 #molecule_type mRNA
 #residues 44-123,'V',125-493 ##label GOO
 #cross-references GB:M58165; NID:9162966; PID:9552318

REFERENCE S38721
 #authors Zimbelmann, R.
 #submission Submitted to the EMBL Data Library, February 1991
 #accession S38721
 #molecule_type mRNA
 #residues 44-1043 ##label ZIM
 #cross-references EMBL:X57784; NID:9436061; PID:9436062

REFERENCE A48173
 #authors Koch, P.J.; Walsh, M.J.; Schmelz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
 #journal Eur. J. Cell Biol. (1990) 53:1-12
 #title Identification of desmoglein, a constitutive desmosomal glycoprotein, as a member of the cadherin family of cell adhesion molecules.

#cross-references M01D:91168965
 #accession A48173
 #molecule_type mRNA
 #residues 44-1001 'AQPSPAR' ##label KO3
 #cross-references GB:X57784
 ##note this sequence has been revised in references A38872 and S38721

GENETICS
 #gene DSG1
 #classification #superfamily cadherin; cadherin repeat homology calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein

FEATURE
 1-23
 24-49
 50-1043
 50-548
 52-157
 160-269
 272-385
 392-491
 549-574
 575-1043
 846-875
 876-905
 906-933
 934-962
 963-1012
 110
 180,496

SUMMARY
 #length 1043 #molecular-weight 112242 #checksum 6897

Query Match 80.0% Score 84; DB 1; Length 1043;
 Best Local Similarity 60.0%; Pred. No. 6.33e-05;
 Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 97 YGIFVINKTGEINI 111
 |||||:|||||
 1 FGIFVVDKNTGDIINI 15

RESULT 3
 ENTRY IJBOGI #type complete
 TITLE desmoglein 1 precursor - human
 ALTERNATE_NAMES desmosomal glycoprotein I
 ORGANISM Homo sapiens #common.name man
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
 26-Feb-1998

ACCESSIONS S16906; A39706; A61254; A61279; S16158
 REFERENCE S16906
 #authors Buxton, R.S.
 #submission Submitted to the EMBL Data Library, November 1990
 #accession S16906
 #molecule_type mRNA
 #residues 1-1049 ##label BUX
 #cross-references EMBL:X56554; NID:930505; PID:930506

REFERENCE A39706
 #authors Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Ataliotis, P.; Poynter, D.; Arneemann, J.; Rutman, A.J.; Pidsley, S.C.; Watt, F.M.; Rees, D.A.; Buxton, R.S.; Magee, A.I.
 #journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:4756-4800
 #title Desmosomal glycoprotein dg1, a component of intercellular desmosome junctions, is related to the cadherin family of cell adhesion molecules.

#cross-references M01D:91271279
 #accession A39706
 #molecule_type mRNA
 #residues 24-1049 ##label WHR
 #cross-references GB:X56654

REFERENCE A61254
 #authors Nilles, L.A.; Parry, D.A.D.; Powers, E.E.; Angst, B.D.; Wagner, R.M.; Green, K.J.
 #journal J. Cell Sci. (1991) 99:809-821
 #title Structural analysis and expression of human desmoglein: a cadherin-like component of the desmosome.

#accession A61254
 #molecule_type mRNA
 #residues 26-1049 ##label NIL

REFERENCE A61279
 #authors Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arneemann, J.; Rees, D.A.; King, I.A.; Magee, A.I.; Blochem. Soc. Trans. (1991) 19:1060-1064
 #journal Desmosomal glycoproteins I, II and III: novel members of the

cadherin superfamily.
#cross-references MUID:92175187
#accession A61279 not compared with conceptual translation
#status not compared with conceptual translation
#molecule-type mRNA
#residues 1-55 #label WH3

GENETICS
#gene GDB:DSG1
#cross-references GDB:125653; OMIM:125670
#map_position 18q12.1-18q12.2
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein

FEATURE
1-23 #domain signal sequence #status predicted #label SIG
24-49 #domain propeptide #status predicted #label PRO
50-1049 #product desmoglein #status predicted #label MAT
50-548 #domain extracellular #status predicted #label EXT
52-157 #domain cadherin repeat homology #label CR1
160-269 #domain cadherin repeat homology #label CR2
272-385 #domain cadherin repeat homology #label CR3
392-493 #domain cadherin repeat homology #label CR4
509-530 #region serine/threonine-rich
549-569 #domain intracellular #status predicted #label TMN
572-1049 #domain desmoglein repeat #label DG1
840-869 #domain desmoglein repeat #label DG2
870-899 #domain desmoglein repeat #label DG3
900-927 #domain desmoglein repeat #label DG4
928-956 #domain desmoglein repeat #label DG4
969-1019 #region glycine/serine-rich
110,180 #binding-site carbohydrate (Asn) (covalent) #status predicted

SUMMARY #length 1049 #molecular-weight 113715 #checksum 4482

Query Match 80.0%; Score 84; DB 1; Length 1049;
Best Local Similarity 60.0%; Pred. No. 6,33e-05;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 97 YGIFVVKDGTGDI 111
OY 1 FGIFVVDKNTGDI 15

RESULT 4
ENTRY #type complete
TITLE desmoglein 2 - human
ALTERNATE_NAMES desmoglein HDGC
ORGANISM #formal_name Homo sapiens #common_name man
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Sep-1997
ACCESSIONS S38673; B38872
REFERENCE S38673
#authors Zimmerman, R.
#submission submitted to the EMBL Data Library, September 1993
#accession S38673
#status preliminary
#molecule-type mRNA
#residues 1-1117 #label ZIM
REFERENCE A38872
#cross-references EMBL:Z26317; NID:9416177; PID:9416178
#authors Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.
#journal Eur. J. Cell Biol. (1991) 55:200-208
#title Complete amino acid sequence of the epidermal desmoglein precursor polypeptide and identification of a second type of desmoglein gene.
#accession B38872
#molecule-type mRNA
#residues 777-1117 #label KOC
GENETICS #cross-references GB:S64273
#gene GDB:DSG2

#cross-references GDB:128808; OMIM:125671
#map_position 18q12.1-18q12.2
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein; membrane protein

FEATURE
51-158 #domain cadherin repeat homology #label CR1
161-271 #domain cadherin repeat homology #label CR2
SUMMARY #length 1117 #molecular-weight 123384 #checksum 7660

Query Match 78.1%; Score 82; DB 2; Length 1117;
Best Local Similarity 60.0%; Pred. No. 1.66e-04;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 98 FGIFVVDKNTGDI 112
OY 1 FGIFVVDKNTGDI 15

RESULT 5
ENTRY #type complete
TITLE desmoglein 1a - bovine
ALTERNATE_NAMES desmoglein HDGC; desmosomal glycoprotein 2
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 24-Oct-1997
ACCESSIONS A43838; A38456; A60714; S14542
REFERENCE A43838
#authors Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Schmelz, M.; Franke, W.W.
#journal Differentiation (1991) 47:29-36
#title Amino acid sequence of bovine muzzle epithelial desmoglein derived from cloned cDNA: a novel subtype of desmosomal cadherins.
#accession A43838
#cross-references MUID:92008912
#molecule-type mRNA
#residues 1-761 #label KOC
#cross-references GB:X58029; GB:S57985; NID:9453583; PID:9453584
#accession B43838
#molecule-type protein
#residues 1-32;65-76;148-159;164-176;190-205;208-219;238-256;
361-375;377-388;478-486 #label KO2
#experimental_source muzzle epithelium
#note sequence extracted from NCBI backbone
REFERENCE A38456
#authors Collins, J.E.; Legan, P.K.; Kenny, T.P.; Macgarvie, J.; Holton, J.L.; Garrido, D.R.
#journal J. Cell Biol. (1991) 113:381-391
#title Cloning and sequence analysis of desmosomal glycoproteins 2 and 3 (desmocolins): cadherin-like desmosomal adhesion molecules with heterogeneous cytoplasmic domains.
#cross-references MUID:91185414
#accession A38456
#molecule-type mRNA
#residues 606-761 #label COL
REFERENCE A60714
#cross-references EMBL:X56967; NID:9310; PID:9311
#authors Holton, J.L.; Kenny, T.P.; Legan, P.K.; Collins, J.E.; Keen, J.N.; Sharma, R.; Garrido, D.R.
#journal J. Cell Sci. (1990) 97:239-246
#title Desmosomal glycoproteins 2 and 3 (desmocolins) show N-terminal similarity to calcium-dependent cell-cell adhesion molecules.
#accession A60714
#molecule-type protein
#residues 1-6,'A',8-9,'R',11-17,'RCE',21-23 #label HOL
#experimental_source nasal epidermis

GENETICS
#gene DSC1
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS alternative splicing; calcium binding; cell adhesion; duplication; glycoprotein; phosphoprotein; transmembrane

```

FEATURE
1-761      #product desmocollin 1a #status experimental #label MAT\
1-561      #domain extracellular #status predicted #label EXT\
3-108      #domain cadherin repeat homology #label CR1\
111-720    #domain cadherin repeat homology #label CR2\
223-338    #domain cadherin repeat homology #label CR3\
339-444    #domain cadherin repeat homology #label CR4\
445-561    #domain cadherin repeat homology #label CR5\
562-582    #domain transmembrane #status predicted #label TM\
583-761    #domain intracellular #status predicted #label INT\
31,266,413 #binding-site carbohydrate (Asn) (covalent) #status
          predicted\
584,588,678 #binding-site phosphate (Thr) (covalent) (by protein
          kinase C) #status predicted\
605         #binding-site phosphate (Ser) (covalent) (by caseln
          kinase II) #status predicted\
671         #binding-site phosphate (Ser) (covalent) (by protein
          kinase C) #status predicted\
681         #binding-site phosphate (Tyr) (covalent) #status
          predicted\
682         #binding-site phosphate (Thr) (covalent) (by caseln
          kinase II) #status predicted\
SUMMARY    #length 761 #molecular-weight 85170 #checksum 3508

Query Match      62.9%; Score 66; DB 1; Length 761;
Best Local Similarity 53.8%; Pred. No. 2,456-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 48 FNFLEKDTGDI 60
1 : ::::|
1 FGIFVVDKNTGDI 13

RESULT 6
ENTRY 150180 #type complete
TITLE cadherin-7 - chicken
ORGANISM #formal_name Gallus gallus #common_name Chicken
DATE 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change
09-May-1997

ACCESSIONS 150178
REFERENCE #authors Nakagawa, S.; Takeichi, M.
          #journal Development (1995) 121:1321-1332
          #title Neural crest cell-cell adhesion controlled by sequential
          #cross-references MIMD:95309115
          #accession 150180
          #status preliminary; translated from GB/EMBL/DBJ
          #molecule_type mRNA
          #residues 1-785 #label NAK
          #cross-references GB:D42150; NID:9868000; PID:9868001
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
SUMMARY #length 785 #molecular-weight 87171 #checksum 8490

Query Match      62.9%; Score 66; DB 2; Length 785;
Best Local Similarity 61.5%; Pred. No. 2,456-01;
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 93 SIFLIDNTGDIH 105
1 : ::::|
2 GIFVVDKNTGDIH 14

RESULT 7
ENTRY 1J0BDF #type complete
TITLE desmocollin 1b precursor - bovine
ALTERNATE_NAMES desmosomal glycoprotein 3
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
24-Oct-1997

ACCESSIONS B38456; A39377; S14567
REFERENCE B38456

```

```

#authors Collins, J.E.; Legan, P.K.; Kenny, T.P.; MacGarvie, J.;
#journal Holton, J.L.; Garrod, D.R.
#title J. Cell Biol. (1991) 113:381-391
          Cloning and sequence analysis of desmosomal glycoproteins 2
          and 3 (desmocollins): cadherin-like desmosomal adhesion
          molecules with heterogeneous cytoplasmic domains.
#cross-references MIMD:91185414
#accession B38456
#molecule_type mRNA
#residues 1-839 #label COL
#cross-references GB:X56966; NID:93315; PID:93316
REFERENCE A39377
#authors Mechanic, S.; Raynor, K.; Hill, J.E.; Cowin, P.
#journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:4476-4480
#title Desmocollins form a distinct subset of the cadherin family of
          cell adhesion molecules.
#cross-references MIMD:91238591
#accession A39377
#molecule_type mRNA
#residues 1-484, 'A', 486-839 #label MEC
#cross-references GB:M67489; GB:M61750; NID:9162970; PID:9162971
#note part of this sequence, including the amino end of the
          mature protein, was confirmed by protein sequencing

GENETICS
#gene DSCI
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS alternative splicing; calcium binding; cell adhesion;
          duplication; glycoprotein; phosphoprotein; transmembrane
          protein

FEATURE
1-29      #domain signal sequence #status predicted #label SIG\
30-132    #domain propeptide #status predicted #label PRO\
133-839   #product desmocollin 1b #status experimental #label MAT\
133-682   #domain extracellular #status predicted #label EXT\
133-240   #domain cadherin repeat homology #label CR1\
243-352   #domain cadherin repeat homology #label CR2\
355-470   #domain cadherin repeat homology #label CR3\
471-576   #domain cadherin repeat homology #label CR4\
577-682   #domain cadherin repeat homology #label CR5\
594-714   #domain cadherin repeat homology #label CR5\
718-839   #domain transmembrane #status predicted #label TM\
163,398,545 #domain intracellular #status predicted #label INT\
          #binding-site carbohydrate (Asn) (covalent) #status
          predicted\
716,720,810 #binding-site phosphate (Thr) (covalent) (by protein
          kinase C) #status predicted\
737        #binding-site phosphate (Ser) (covalent) (by caseln
          kinase II) #status predicted\
803,830    #binding-site phosphate (Ser) (covalent) (by protein
          kinase C) #status predicted\
813        #binding-site phosphate (Tyr) (covalent) #status
          predicted\
814        #binding-site phosphate (Thr) (covalent) (by caseln
          kinase II) #status predicted\
SUMMARY #length 839 #molecular-weight 93521 #checksum 6365

Query Match      62.9%; Score 66; DB 1; Length 839;
Best Local Similarity 53.8%; Pred. No. 2,456-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 180 FNFLEKDTGDI 192
1 : ::::|
1 FGIFVVDKNTGDI 13

RESULT 8
ENTRY B48910 #type fragment
TITLE desmocollin 1b precursor - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 08-Dec-1995 #sequence_revision 08-Dec-1995 #text_change
15-Aug-1997

ACCESSIONS B48910
REFERENCE A48910
          King, I.A.; Arnemann, J.; Spurr, N.K.; Buxton, R.S.

```

#journal Genomics (1993) 18:185-194
#title Cloning of the cDNA (DSCI) coding for human type 1
#cross-references MUID:94116981
#accession B48910
#status preliminary
#molecule_type mRNA
#residues 1-770 ##label KIN
#cross-references GB:X72925
GENETICS
#gene GDB:DSCI
#cross-references GDB:128632
#map_position 18q12.2-18q12.2
CLASSIFICATION
#superfamily cadherin; cadherin repeat homology
KEYWORDS
alternative splicing; calcium binding; cell adhesion;
glycoprotein; phosphoprotein; transmembrane protein
FEATURE
67-172 #domain cadherin repeat homology #label CRI
SUMMARY
#length 770 #checksum 4861
Query Match 61.9%; Score 65; DB 2; Length 770;
Best Local Similarity 53.8%; Pred. No. 3.76e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
Db 112 FNLFYIEKDTGDI 124
OY 1 FGIFVVDKNTGDI 13
RESULT 9
ENTRY A48910 #type fragment
TITLE desmocollin 1a precursor - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 08-Dec-1995 #sequence_revision 08-Dec-1995 #text_change
15-Aug-1997
ACCESSIONS
#journal A48910
#authors King, I.A.; Arnemann, J.; Spurr, N.K.; Buxton, R.S.
#journal Genomics (1993) 18:185-194
#title Cloning of the cDNA (DSCI) coding for human type 1
#cross-references MUID:94116981
#accession A48910
#status preliminary
#molecule_type mRNA
#residues 1-824 ##label KIN
#cross-references GB:X72925
GENETICS
#gene GDB:DSCI
#cross-references GDB:128632
#map_position 18q12.2-18q12.2
CLASSIFICATION
#superfamily cadherin; cadherin repeat homology
KEYWORDS
alternative splicing; calcium binding; cell adhesion;
glycoprotein; phosphoprotein; transmembrane protein
FEATURE
67-172 #domain cadherin repeat homology #label CRI
SUMMARY
#length 824 #checksum 2645
Query Match 61.9%; Score 65; DB 2; Length 824;
Best Local Similarity 53.8%; Pred. No. 3.76e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
Db 112 FNLFYIEKDTGDI 124
OY 1 FGIFVVDKNTGDI 13
RESULT 10
ENTRY I37281 #type complete
TITLE Dscia precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
16-Feb-1997

ACCESSIONS
#journal I37281
#authors Theis, D.G.; Koch, P.J.; Franke, W.W.
#journal Int. J. Dev. Biol. (1993) 37:101-110
#title Differential synthesis of type 1 and type 2 desmocollin mRNAs
#cross-references MUID:93283249
#accession I37281
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-840 ##label RES
#cross-references EMBL:Z34522; NID:9505536; PID:9505537
CLASSIFICATION
#superfamily cadherin; cadherin repeat homology
SUMMARY
#length 840 #molecular_weight 93848 #checksum 3174
Query Match 61.9%; Score 65; DB 2; Length 840;
Best Local Similarity 53.8%; Pred. No. 3.76e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
Db 182 FNLFYIEKDTGDI 194
OY 1 FGIFVVDKNTGDI 13
RESULT 11
ENTRY I37282 #type complete
TITLE Dscib precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
16-Feb-1997
ACCESSIONS
#journal I37282
#authors Theis, D.G.; Koch, P.J.; Franke, W.W.
#journal Int. J. Dev. Biol. (1993) 37:101-110
#title Differential synthesis of type 1 and type 2 desmocollin mRNAs
#cross-references MUID:93283249
#accession I37282
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-894 ##label RES
#cross-references EMBL:Z34522; NID:9505536; PID:9505538
GENETICS
#incons 829/2
CLASSIFICATION
#superfamily cadherin; cadherin repeat homology
SUMMARY
#length 894 #molecular_weight 100044 #checksum 7946
Query Match 61.9%; Score 65; DB 2; Length 894;
Best Local Similarity 53.8%; Pred. No. 3.76e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
Db 182 FNLFYIEKDTGDI 194
OY 1 FGIFVVDKNTGDI 13
RESULT 12
ENTRY S59969 #type complete
TITLE DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) isoform
beta - Chinese hamster
ORGANISM #formal_name Crictetus griseus #common_name Chinese hamster
DATE 15-Feb-1996 #sequence_revision 13-Mar-1997 #text_change
17-Mar-1999
ACCESSIONS
#journal S59969
#authors Derudder, S.; Frey, S.; Delaporte, C.; Jacquemin-Sablon, A.
#journal Biochim. Biophys. Acta (1995) 1264:178-182
#title Cloning and characterization of full-length cDNAs coding for
the DNA topoisomerase II beta from Chinese hamster lung
cells sensitive and resistant to 9-OH-ellipticine.
#cross-references MUID:96085121
#accession S59969

##status translation not shown
##molecule_type mRNA
##residues 1-1612 ##label DER
##cross-references EMBL:X86455; NID:g790987; PID:g790988
##experimental_source lung
CLASSIFICATION #superfamily eukaryotic type II DNA topoisomerase; phage T4
DNA topoisomerase (ATP-hydrolyzing) medium chain homology
KEYWORDS ATP; DNA binding; DNA replication; isomerase; nucleus
FEATURES
697-927
SUMMARY #domain phage T4 DNA topoisomerase (ATP-hydrolyzing)
medium chain homology #label TOP
#length 1612 #molecular-weight 182074 #checksum 5332
Query Match 58.1%; Score 61; DB 2; Length 1612;
Best Local Similarity 72.7%; Pred. No. 2.01e+00;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 922 IFVVDKNTVEI 932
QY 3 IFVVDKNTGDI 13

RESULT 13
ENTRY A39242 #type complete
TITLE DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) beta,
splice form 2 - human
ALTERNATE_NAMES DNA topoisomerase II isoform beta-2
CONTAINS DNA topoisomerase II isoform beta-1
ORGANISM #formal_name Homo sapiens #common_name man
DATE 04-Oct-1991 #sequence_revision 03-May-1996 #text_change
17-Mar-1999
ACCESSIONS 526730; A39242; S10710; S33970; S30191; S41641; S30190
REFERENCE 526730
#authors Jenkins, J.R.; Ayton, P.; Jones, S.L.; Simmons,
D.L.; Harris, A.L.; Sheer, D.; Hickson, I.D.
#journal Nucleic Acids Res. (1992) 20:5587-5592
#title Isolation of cDNA clones encoding the beta isozyme of human
DNA topoisomerase II and localisation of the gene to
chromosome 3p24.
#cross-references MUID:93087165
#accession 526730
##molecule_type mRNA
##residues 1-23,29-1626 ##label JEN
##cross-references EMBL:X68060; NID:g37230; PID:g37231
REFERENCE A39242
#authors Chung, T.D.-Y.; Drake, F.H.; Tan, K.B.; Per, S.R.; Crooke,
S.T.; Mirabelli, C.K.
#journal Proc. Natl. Acad. Sci. U.S.A. (1989) 86:9431-9435
#title Characterization and immunological identification of cDNA
clones encoding two human DNA topoisomerase II isozymes.
#cross-references MUID:90083281
#accession A39242
##molecule_type mRNA
##residues 149-1043 ##label CHU
##cross-references GB:M27504
REFERENCE S10710
#authors Austin, C.A.; Fisher, L.M.
#journal FEBS Lett. (1990) 266:115-117
#title Isolation and characterization of a human cDNA clone encoding
a novel DNA topoisomerase II homologue from HeLa cells.
#cross-references MUID:90306333
#accession S10710
##molecule_type mRNA
##residues 1043-1276 ##label AUS
##cross-references GB:X53662; GB:S56813; NID:g38324; PID:g38325
REFERENCE S30190
#authors Austin, C.A.; Sng, J.H.; Patel, S.; Fisher, L.M.
#journal Biochim. Biophys. Acta (1993) 1172:283-291
#title Novel HeLa topoisomerase II is the II-beta isoform: complete
coding sequence and homology with other type II
topoisomerases.
#cross-references MUID:93192319
#accession S33970

##molecule_type mRNA
##residues 1-23,29-1610,'A',1612-1626 ##label AUV
##cross-references EMBL:Z15111
REFERENCE S30191
#authors Austin, C.A.; Sng, J.H.; Patel, S.; Fisher, L.M.
#submission submitted to the EMBL Data Library, September 1992
#accession S30191
##molecule_type mRNA
##residues 596-1430,'S',1432-1610,'A',1612-1626 ##label AUV
##cross-references EMBL:Z15115; NID:g288564; PID:g288565
REFERENCE S41641
#authors Davies, S.L.; Jenkins, J.R.; Hickson, I.D.
#journal Nucleic Acids Res. (1993) 21:3719-3723
#title Human cells express two differentially spliced forms of
topoisomerase II-beta mRNA.
#cross-references MUID:93376494
#accession S41641
##molecule_type DNA
##residues 24-80 ##label DAV
##cross-references EMBL:X71911
##note this sequence represents a long minor splice form,
designated beta-2

GENETICS
#gene GDB:TOP2B
#cross-references GDB:131575; OMIM:126431
#map_position 3p24-3p24
CLASSIFICATION #superfamily eukaryotic type II DNA topoisomerase; phage T4
DNA topoisomerase (ATP-hydrolyzing) medium chain homology
alternative splicing; ATP; dimer; isomerase; nucleus
KEYWORDS
FEATURES
1-1626
#product DNA topoisomerase II beta-2 #status predicted
#label MINR
#product DNA topoisomerase II beta-1 #status predicted
#label MAJR
SUMMARY #domain phage T4 DNA topoisomerase (ATP-hydrolyzing)
medium chain homology #label TOP #checksum 3565
#length 1626 #molecular-weight 183296

Query Match 58.1%; Score 61; DB 2; Length 1626;
Best Local Similarity 72.7%; Pred. No. 2.01e+00;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db .934 IFVVDKNTVEI 944
QY 3 IFVVDKNTGDI 13

RESULT 14
ENTRY A65141 #type complete
TITLE hypothetical 10.6 kD protein in gntR-ggt intergenic region -
Escherichia coli (strain K-12)
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
14-Nov-1997
ACCESSIONS A65141
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Colado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession A65141
##status preliminary; nucleic acid sequence not shown;
translation not shown

##molecule_type DNA
##residues 1-94 ##label BLAT
##cross-references GB:AE000421; GB:U00096; NID:g1789854; PID:g1789855;
#accession UMGCP.D3446

##experimental_source strain K-12, substrain MG1655

GENETICS

```

#gene      yf1B      #length 94      #molecular-weight 10613      #checksum 4152
SUMMARY

Query Match      57.1%      Score 60; DB 2; Length 94;
Best Local Similarity 45.5%      Pred. No. 3.02e+00;
Matches 5; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Db      59      FIIDKSGEIH 69
      ||::||::||::||:
Oy      4      FVVDKNTGDIN 14

RESULT 15
ENTRY      S76379      #type complete
TITLE      hypothetical protein - Synechocystis sp. (strain PCC 6803)
ORGANISM   PCC 6803
VARIETY    #variety
DATE       25-Apr-1997 #sequence-revision 25-Apr-1997 #text-change
          21-Aug-1998
ACCESSIONS S76379
REFERENCE  S74322
AUTHORS     Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
            Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugita, M.;
            Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
            Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpou,
            S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
            Yasuda, M.; Tabata, S.
            DNA Res. (1996) 3:109-136
            Sequence analysis of the genome of the unicellular
            cyanobacterium Synechocystis sp. PCC6803. II. Sequence
            determination of the entire genome and assignment of
            potential protein-coding regions.
            #cross-references MIMD:97061201
            #accession S76379
            #status preliminary
            #molecule-type DNA
            ##residues 1-214 ##label KAN
            #cross-references EMBL:D64000; GB:AB001339; NID:g1001484; PID:d1010882;
            PID:g1001603
            #note
            the nucleotide sequence was submitted to the EMBL Data
            Library, June 1996
            #length 214 #molecular-weight 24539 #checksum 4859
SUMMARY

Query Match      57.1%      Score 60; DB 2; Length 214;
Best Local Similarity 42.9%      Pred. No. 3.02e+00;
Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db      49      FGVMIDPSTGEIS 62
      ||::||::||::||:
Oy      1      FGIFVVDKNTGDIN 14

```

Search completed: Fri Jun 11 17:20:20 1999
Job time : 86 secs.

(TM)

Distribution rights by Oxford Molecular Ltd

generated

1 EGI FVVDKNTGDINI 15

Cap 15

Listing first 45 summaries

1:swissprot

Mean 27.494; Variance 34.835; scale 0.789

ved by analysis of the total score distribution.

SUMMARIES

3.3 756 1 YD34_SCHPO HYPOTHETICAL 81.7 KD P 5.10e+00

45	53	50.5	794	1	CADC_HUMAN	BRAIN-CADHERIN PRECURS	1.82e+0
----	----	------	-----	---	------------	------------------------	---------

ALIGNMENTS

RESULT	1	STANDARD:	PR:	999 AA.
ID	DSG3	HUMAN		
AC	P32926:			
DT	01-OCT-1993	(REL. 27, CREATED)		
DT	01-OCT-1993	(REL. 27, LAST SEQUENCE UPDATE)		
DT	01-NOV-1997	(REL. 35, LAST ANNOTATION UPDATE)		
DE	DESMOGLEIN 3	PREDICATOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).		
GN	DSG3.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUMECROTIA, METACOTA: CHORDATA: VERTEBRATA: MAMMALIA: EUTHERIA:			
CC	PRIMATES: CATARRHINI: HOMINIDAE: HOMO.			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 92069753.			
RA	AMAGAI M., KLAUS-KOYTUN V., STANLEY J.R.:			
RT	"Antibodies against a novel epithelial cadherin in pemphigus			
RL	vulgaris, a disease of cell adhesion."			
CC	CELL 67:869-877(1991).			
CC	-1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.			
CC	INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE			
CC	FILAMENTS MEDIATING CELL-CELL ADHESION.			
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.			
CC	-1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, OESOPHAGUS AND			
CC	CARCINOMAS.			
CC	-1- DOMAIN: CALCULIN MAY BE BOUND BY THE CADHERIN-LIKE REPEATS			
CC	(POTENTIAL).			
CC	-1- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN			
CC	DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE			
CC	LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES			
CC	AGAINST DSG3.			
CC	-1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE			
CC	DESMOSOMAL SUBFAMILY.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: M76482; GI90752; -			
DR	PIR: A41088; ITHUG3.			
DR	MIM: 169615; -			
DR	PROSITE: PS00232; CADHERIN. 3.			
DR	PFAM: PF00026; cadherin. 4.			
DR	HSSP: P09803; IEDH.			

KW CELL ADHESION; SIGNAL; TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
KW CALCIUM-BINDING; REPEAT.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 49 POTENTIAL.
FT CHAIN 50 999 DESMOGLEIN 3.
FT DOMAIN 50 615 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 616 640 POTENTIAL.
FT DOMAIN 641 999 CYTOPLASMIC (POTENTIAL).
FT REPEAT 50 158 CADHERIN 1.
FT REPEAT 159 268 CADHERIN 2.
FT REPEAT 269 383 CADHERIN 3.
FT REPEAT 386 499 CADHERIN 4.
FT REPEAT 910 935 DESMOGLEIN REPEAT 1.
FT REPEAT 936 966 DESMOGLEIN REPEAT 2.
FT REPEAT 110 110 POTENTIAL.
FT CARBOHYD 180 180 POTENTIAL.
FT CARBOHYD 459 459 POTENTIAL.
FT CARBOHYD 545 545 POTENTIAL.
SQ SEQUENCE 999 AA; 107503 MW; 4891F6AE CRC32;
Query Match 100.0%; Score 105; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 2,66e-11;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 97 FGIFVYDKNTGDINI 111
QY 1 FGIFVYDKNTGDINI 15
RESULT 2
ID DSG1 BOVIN STANDARD; PRT; 1043 AA.
AC 003763;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DG1).
GN DSG1
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDEA; BOVINAE; BOS.
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-MUZZLE EPITHELIUM;
RA KOCH P.J., GOLDSCHMIDT M.D., ZIMBELMANN R., FRANK W.W.;
RL SUBMITTED (MAR-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
RN (2)
RP SEQUENCE OF 44-1043 FROM N.A.
RC TISSUE-MUZZLE EPITHELIUM;
RA KOCH P.J., WALSH M.J., SCHMELZ M., GOLDSCHMIDT M.D.,
RA ZIMBELMANN R., FRANK W.W.;
RT Identification of desmoglein, a constitutive desmosomal
RT glycoprotein, as a member of the cadherin family of cell adhesion
RT molecules."
RL EUR. J. CELL BIOL. 53:1-12(1990).
RN (3)
RP REVISIONS, AND SEQUENCE OF 101-123.
RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R.,
RA FRANK W.W.;
RT "Complete amino acid sequence of the epidermal desmoglein precursor
RT polypeptide and identification of a second type of desmoglein gene."
RL EUR. J. CELL BIOL. 55:200-208(1991).
RN (4)
RP SEQUENCE OF 44-493 FROM N.A.
RA MEDLINE; 91097553.
RA GOODMAN L., HILL J.E., RAYNOR K., RASZI L., MANABE M., COWIN P.;
RT "Desmoglein shows extensive homology to the cadherin family of cell
RT adhesion molecules."
RL BIOCHEM. BIOPHYS. RES. COMMUN. 173:1224-1230(1990).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EPIDERMIS, MUZZLE, TONGUE AND OESOPHAGUS.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE
CC DESMOsome SUBFAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X58466; G307; -
DR EMBL; X57784; G436062; -
DR EMBL; M58165; G552318; -
DR PIR; S14603; I1B0G1.
DR PROSITE; PS00232; CADHERIN; 2.
DR PFAM; PF00028; cadherin; 3.
DR HSSP; P09803; 1BDH.
DR CELL ADHESION; SIGNAL; TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
KW CALCIUM-BINDING; REPEAT.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 49 POTENTIAL.
FT CHAIN 50 1043 DESMOGLEIN 1.
FT DOMAIN 50 548 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 549 573 POTENTIAL.
FT DOMAIN 574 1043 CYTOPLASMIC (POTENTIAL).
FT REPEAT 50 158 CADHERIN 1.
FT REPEAT 159 270 CADHERIN 2.
FT REPEAT 271 385 CADHERIN 3.
FT REPEAT 386 498 CADHERIN 4.
FT REPEAT 819 845 DESMOGLEIN REPEAT 1.
FT REPEAT 846 875 DESMOGLEIN REPEAT 2.
FT REPEAT 876 905 DESMOGLEIN REPEAT 3.
FT REPEAT 906 933 DESMOGLEIN REPEAT 4.
FT REPEAT 934 962 DESMOGLEIN REPEAT 5.
FT DOMAIN 963 1012 GLY/SER-RICH.
FT CARBOHYD 110 110 POTENTIAL.
FT CARBOHYD 180 180 POTENTIAL.
FT CARBOHYD 496 496 POTENTIAL.
FT CONFLICT 124 124 I -> V (IN REF. 4).
SQ SEQUENCE 1043 AA; 112243 MW; 13698584 CRC32;
Query Match 80.0%; Score 84; DB 1; Length 1043;
Best Local Similarity 60.0%; Pred. No. 4.50e-06;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 97 YGIFVYDKNTGDINI 111
QY 1 FGIFVYDKNTGDINI 15
RESULT 3
ID DSG1 HUMAN STANDARD; PRT; 1049 AA.
AC Q02413;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DG1).
GN DSG1
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-KERATINOCYTES;
RA WHEELER G.N., PARKER A.E., THOMAS C.L., ATALLOTIS P., POYNTER D.,
RA ARSEMAN J., RUTMAN A.J., PIDSLEY S.C., WATT F.M., REES D.A.,
RA BOXTON R.S., MAEE A.I.;

RT "Desmosomal glycoprotein Dsg2, a component of intercellular desmosome junctions, is related to the cadherin family of cell adhesion molecules.";
RT PROC. NATL. ACAD. SCI. U.S.A. 88:4796-4800(1991).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENT MEDIANING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSILLAND OESOPHAGUS.
CC -1- DOMAIN: CALCULUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE DESMOSOMAL SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: X56654; G30506; .
DR PIR: S16906; IJHUG1.
DR MIM: 125670; .
DR PROSITE: PS00232; CADHERIN; 2.
DR PFM: PF00028; cadherin; 4.
DR HSSP: P09803; IEDH.
KW CELL ADHESION; SIGNAL; TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
KW CALCIUM-BINDING; REPEAT.
FT SIGNAL 1 23
FT PROPEP 24 49
FT CHAIN 50 1049
FT DOMAIN 50 545
FT TRANSMEM 546 570
FT DOMAIN 571 1049
FT REPEAT 50 158
FT REPEAT 159 270
FT REPEAT 271 385
FT REPEAT 386 497
FT REPEAT 497 839
FT REPEAT 840 869
FT REPEAT 870 899
FT REPEAT 900 927
FT REPEAT 928 956
FT DOMAIN 969 1019
FT CARBOHYD 36 36
FT CARBOHYD 110 110
FT CARBOHYD 180 180
SQ SEQUENCE 1049 AA; FDD79961 CRC32;
Query Match 80.0%; Score 84; DB 1; Length 1049;
Best Local Similarity 60.0%; Pred. No. 4.50e-06;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 97 YGIVFNKDTGELN 111
QY 1 FGIVFNKDTGDI 15
RESULT 4
ID DSG2_HUMAN STANDARD: PRT; 1117 AA.
AC Q14126;
DT 01-NOV-1997 (REL. 35; CREATED)
DT 01-NOV-1997 (REL. 35; LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35; LAST ANNOTATION UPDATE)
DE DESMOGLEIN 2 PRECURSOR (HDGC).
GN DSG2.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.

RC TISSUE-COLON CARCINOMA;
RX MEDLINE: 94192736.
RA SCHAEFER S., KOCH P.J., FRANK W.W.;
RT "Identification of the ubiquitous human desmoglein, Dsg2, and the expression catalogue of the desmoglein subfamily of desmosomal cadherins.";
RT EXP. CELL RES. 211:391-399(1994).
RN [2]
RP SEQUENCE OF 777-1117 FROM N.A.
RX MEDLINE: 92037656.
RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMMERMANN R., FRANK W.W.;
RT "Complete amino acid sequence of the epidermal desmoglein precursor polypeptide and identification of a second type of desmoglein gene.";
RT EUR. J. CELL BIOL. 55:200-208(1991).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENT MEDIANING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.
CC -1- DOMAIN: CALCULUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE DESMOSOMAL SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: 226317; G416178; .
DR MIM: 125671; .
DR PROSITE: PS00232; CADHERIN; 3.
DR PFM: PF00028; cadherin; 4.
DR HSSP: P15116; INCI.
KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;
KW CYTOSKELETON; CALCIUM-BINDING.
FT SIGNAL 1 23
FT PROPEP 24 48
FT CHAIN 49 1117
FT DOMAIN 49 608
FT TRANSMEM 609 633
FT DOMAIN 634 1117
FT REPEAT 634 159
FT REPEAT 160 272
FT REPEAT 273 387
FT REPEAT 388 502
FT REPEAT 502 911
FT REPEAT 912 941
FT REPEAT 942 967
FT REPEAT 968 991
FT REPEAT 992 1020
FT REPEAT 1021 1050
FT CARBOHYD 111 111
FT CARBOHYD 181 181
FT CARBOHYD 308 308
FT CARBOHYD 461 461
FT CARBOHYD 513 513
SQ SEQUENCE 1117 AA; 122385 MW; 84D3B898 CRC32;
Query Match 78.1%; Score 82; DB 1; Length 1117;
Best Local Similarity 60.0%; Pred. No. 1.34e-05;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 98 FGIVFNKDTGELN 112
QY 1 FGIVFNKDTGDI 15
RESULT 5
ID DSG1_BOVIN STANDARD: PRT; 893 AA.

AC Q01107; Q28095; 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE DESMOCOLLIN 1A/1B PRECURSOR (DESMOSOMAL GLYCOPROTEIN 2/3) (DG2/DG3).
 GN DSC1.
 OS BOS TAURUS (BOVINE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 CC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDEA; BOVIDAE; BOVINAE; BOS.
 RN [1]
 RP SEQUENCE FROM N.A. (1B), AND PARTIAL SEQUENCE.
 RC TISSUE-EPIDERMIS;
 RX COLLINS J.E., LEGAN P.K., KENNY T.P., MACGARVIE J., HOLTON J.L.,
 RA GAROD D.R.;
 RT "Cloning and sequence analysis of desmosomal glycoproteins 2 and 3
 (desmocollins): cadherin-like desmosomal adhesion molecules with
 RT heterogeneous cytoplasmic domains.";
 RL J. CELL BIOL. 113:381-391(1991).
 RN [2]
 RP SEQUENCE FROM N.A. (1B), AND PARTIAL SEQUENCE.
 RX MECHANIC S., RAYNOR K., HILL J.E., COMIN P.;
 RT "Desmocollins form a distinct subset of the cadherin family of cell
 RT adhesion molecules.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:4476-4480(1991).
 RN [3]
 RP SEQUENCE OF 133-893 FROM N.A. (1A), AND PARTIAL SEQUENCE.
 RC TISSUE-MUZZLE EPITHELIUM;
 RX MEDLINE; 92008912.
 RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R., SCHWELZ M.,
 RA FRANK W.W.;
 RT "Amino acid sequence of bovine muzzle epithelial desmocollin derived
 RT from cloned cDNA: a novel subtype of desmosomal cadherins.";
 RL DIFFERENTIATION 47:29-36(1991).
 RN [4]
 RP SEQUENCE OF 133-155.
 RX MEDLINE; 9115997.
 RA HOLTON J.L., KENNY T.P., LEGAN P.K., COLLINS J.E., KEEN J.N.,
 RA SHARMA R., GAROD D.R.;
 RT "Desmosomal glycoproteins 2 and 3 (desmocollins) show N-terminal
 RT similarity to calcium-dependent cell-cell adhesion molecules.";
 RL J. CELL SCI. 97:239-246(1990).
 RN [5]
 RP PHOSPHORYLATION.
 RX MEDLINE; 91009551.
 RA PARRISH E.P., MARSTON J.E., MATTEY D.L., MEASURES H.R., VENNING R.,
 RA GAROD D.R.;
 RT "Size heterogeneity, phosphorylation and transmembrane organisation
 RT of desmosomal glycoproteins 2 and 3 (desmocollins) in mock cells.";
 RL J. CELL SCI. 96:239-248(1990).
 CC -1- FUNCTION. COMPONENT OF INTERCELLULAR DESMOSE JUNCTIONS. INVOLVED
 CC IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
 CC MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
 CC POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
 CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS. LINKED
 CC TO THE KERATINIZATION OF EPITHELIAL TISSUES.
 CC -1- SUBCELLULAR LOCATION. TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY. EPIDERMIS AND WRAKIN IN TONGUE PAPILLAE.
 CC -1- DOMAIN. CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -1- PTM. FORM 1A IS PHOSPHORYLATED ON A SERINE BUT FORM 1B IS NOT.
 CC -1- ALTERNATIVE PRODUCTS. TWO FORMS; 1A OR DG2 (SHOWN HERE) AND 1B OR
 CC DG3; ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
 CC -1- SIMILARITY. BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE
 CC DESMOSONAL SUBFAMILY.

CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M67489; G162971; -
 DR EMBL; X56966; G316; -
 DR EMBL; X56967; G311; -
 DR EMBL; X56968; G314; -
 DR EMBL; X56968; G313; -
 DR EMBL; X58029; G453584; -
 DR PIR; B38456; IJBDF.
 DR PIR; A43838; IJBDE.
 DR PIR; A39377; A39377.
 DR PROSITE; PS00232; CADHERIN; 2.
 DR PFAM; PF00028; cadherin; 4.
 DR HSSP; P09803; IEDH.
 KW CELL ADHESION; GLYCOPROTEIN; PHOSPHORYLATION; TRANSMEMBRANE; SIGNAL;
 KW REPEAT; CYTOSKELETON; CALCIUM-BINDING; ALTERNATIVE SPLICING.
 FT SIGNAL 1 29
 FT PROPEP 30 132
 FT CHAIN 133 893
 FT DOMAIN 133 692
 FT TRANSMEM 693 715
 FT DOMAIN 716 893
 FT REPEAT 133 240
 FT REPEAT 241 352
 FT REPEAT 353 470
 FT REPEAT 471 574
 FT REPEAT 575 682
 FT CARBOHYD 163 163
 FT CARBOHYD 398 398
 FT CARBOHYD 545 545
 FT VARIANT 519 519
 FT VARIANT 788 788
 FT VARSPLIC 829 839
 FT VARSPLIC 840 893
 FT CONFLICT 485 485
 SQ SEQUENCE 893 AA; 99647 MW; FF5DD78B CRC32;
 Query Match 62.9%; Score 66; DB 1; Length 893;
 Best Local Similarity 53.8%; Pred. No. 5,11e-02;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Db 180 FNLFIEKDTGDI 192
 Qy 1 FGIFVYDKNTGDI 13
 RESULT 6
 ID DSC1_HUMAN STANDARD; PRT; 894 AA.
 AC Q08554;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE DESMOCOLLIN 1A/1B PRECURSOR (DESMOSOMAL GLYCOPROTEIN 2/3) (DG2/DG3).
 GN DSC1.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 CC PRIMATES; CARNIVORINIA; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-FORESKIN;
 RX MEDLINE; 93283249.
 RA THEIS D.G., KOCH P.J., FRANK W.W.;
 RT "Differential synthesis of type 1 and type 2 desmocollin mRNAs in
 RT human stratified epithelia.";
 RL INT. J. DEV. BIOL. 37:101-110(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-FORESKIN;
 RA ZIMBELMANN R.;
 RL SUBMITTED (JUN-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.

RC TISSUE-SKIN.
 RA KING I.A., ARNEMANN J., SPURR N.K., BUXTON R.S.;
 RT "Cloning of the CDNA (DSCI) coding for human type 1 desmocollin and
 its assignment to chromosome 18.";
 RL GENOMICS 18:185-194(1993).
 RP [4]
 RP SEQUENCE OF 135-151 AND 283-292.
 RA MEDLINE: 94116981.
 RA KING I.A., MAGEE A.I., REES D.A., BUXTON R.S.;
 RT "Keratinization is associated with the expression of a new protein
 related to the desmosomal cadherins Dgit/Iti.";
 RL FEBS LETT 286:9-12(1991).
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
 IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
 CC MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
 CC POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
 CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS. LINKED
 CC TO THE KERATINIZATION OF EPITHELIAL TISSUES.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN EPIDERMIS, LESS IN LYMPH
 CC NODE AND TONGUE.
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, 1A OR DG2 (SHOWN HERE) AND 1B OR
 CC DG3; ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE
 CC DESMOSOMAL SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC -----
 CC EMBL: 234522; G50538; -
 CC EMBL: 234522; G50538; -
 CC EMBL: X72925; G457463; -
 CC EMBL: X72925; G457464; -
 CC MIM: 125643; -
 CC PROSITE: PS00232; CADHERIN; 3.
 CC PFAM: PF00028; cadherin; 5.
 CC HSSP: P09803; 1EDH.
 CC CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;
 CC CYTOSKELETON; CALCIUM-BINDING; ALTERNATIVE SPLICING.
 CC KW SIGNAL. 1 29
 CC FT PROPEP 30 134
 CC FT CHAIN 135 894
 CC FT DOMAIN 135 691
 CC FT TRANSMEM 692 714
 CC FT DOMAIN 715 894
 CC FT REPEAT 135 242
 CC FT REPEAT 243 354
 CC FT REPEAT 355 471
 CC FT REPEAT 472 575
 CC FT REPEAT 576 682
 CC FT CARBOHYD 165 165
 CC FT CARBOHYD 546 546
 CC FT CARBOHYD 830 840
 CC FT VARSPLIC 841 894
 CC FT CONFLICT 132 132
 CC FT SEQUENCE 894 AA; 100044 MW; A287BCA9 CMC32;
 CC
 CC Query Match 61.9%; Score 65; DB 1; Length 894;
 CC Best Local Similarity 53.8%; Pred. No. 8.27e-02;
 CC Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

RESULT 7
 ID DSCI.MOUSE STANDARD; PRT; 886 AA.
 AC P5849;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE DESMOCOLLIN 1A/1B PRECURSOR.
 GN DSCI.
 OS MUS MUSCULUS (MOUSE).
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 CC RODENTIA; SCIROGNATHI; MURIDAE; MURINAE; MUS.
 RP [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57BL/6; TISSUE-SKIN;
 RX MEDLINE: 96420658.
 RA KING I.A., O'BRIEN T.J., BUXTON R.S.;
 RT "Expression of the 'skin-type' desmosomal cadherin DSCI is closely
 RT linked to the keratinization of epithelial tissues during mouse
 RT development.";
 CC J. INVEST. DERMATOL. 107:531-538(1996).
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
 CC IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
 CC MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
 CC POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
 CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS. LINKED
 CC TO THE KERATINIZATION OF EPITHELIAL TISSUES.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -1- ALTERNATIVE PRODUCTS: FORMS 1A AND 1B ARE PRODUCED BY ALTERNATIVE
 CC SPLICING OF THE SAME GENE.
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE
 CC DESMOSOMAL SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC -----
 CC EMBL: X97986; E246528; -
 CC EMBL: X97986; E246529; -
 CC MGI: MGI:109173; DSCI.
 CC PROSITE: PS00232; CADHERIN; 2.
 CC PFAM: PF00028; cadherin; 5.
 CC HSSP: P09803; 1EDH.
 CC CELL ADHESION; SIGNAL; TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
 CC CALCIUM-BINDING; REPEAT; ALTERNATIVE SPLICING.
 CC KW SIGNAL. 1 29
 CC FT PROPEP 30 134
 CC FT CHAIN 135 886
 CC FT DOMAIN 135 691
 CC FT TRANSMEM 692 714
 CC FT DOMAIN 715 886
 CC FT REPEAT 135 242
 CC FT REPEAT 243 354
 CC FT REPEAT 355 471
 CC FT REPEAT 472 575
 CC FT REPEAT 576 682
 CC FT CARBOHYD 165 165
 CC FT CARBOHYD 546 546
 CC FT CARBOHYD 830 840
 CC FT VARSPLIC 840 886
 CC FT SEQUENCE 886 AA; 98953 MW; 6C3C94BA CMC32;
 CC
 CC Query Match 58.1%; Score 61; DB 1; Length 886;
 CC Best Local Similarity 46.2%; Pred. No. 5.44e-01;

```

Matches      6:  Conservative      5:  Mismatches      2:  Indels      0:  Gaps      0:
Db           182 YNLFYKDTGDI 194
OY           1 FGFVVDKNTGDI 13

RESULT 8
ID      TP2B-CRILLO      STANDARD:      PRT: 1612 AA.
AC      064399:
DT      15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DI      15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE      15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DN      DNA TOPOISOMERASE II, BETA ISOZYME (EC 5.99.1.3).
GN      TOP2B.
OS      CRICETULUS LONGICAUDATUS (LONG-TAILED HAMSTER) (CHINESE HAMSTER).
OC      EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC      RODENTIA; SCIROGNATHI; MURIDAE; CRICETINAE; CRICETULUS.
RN      (1)
RP      SEQUENCE FROM N.A.
RC      TISSUE=LUNG;
RX      MEDLINE: 96085121.
RT      DERUDRE S., FREY S., DELAPORTE C., JACQUEMIN-SABON A.;
RT      Cloning and characterization of full-length cDNAs coding for the DNA
RT      topoisomerase II beta from Chinese hamster lung cells sensitive and
RT      resistant 9-OH-ethylapicline.
RL      BIOCIN. BIOPHYS. ACTA 1264:178-182(1995).
CC      -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC      BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC      MAKES DOUBLE-STRAND BREAKS.
CC      -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC      OF DOUBLE-STRANDED DNA.
CC      -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
CC      -1- EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH NEGATIVE AND
CC      POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES RELAX ONLY
CC      NEGATIVE SUPERCOILS.
CC      -1- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: X86455; G790988;
DR      PROSITE: PS00177; TOPOISOMERASE-II; 1.
DR      PFAM: PF00204; DNA_topoisom; 1.
DR      HSSP: P06786; 1BGW.
KW      ISOMERASE; TOPOISOMERASE; DNA-BINDING; ATP-BINDING; NUCLEAR PROTEIN.
FT      ACT SITE 814 814 DNA CLEAVAGE (BY SIMILARITY).
SQ      SEQUENCE 1612 AA; 182074 MW; AA9562E7 CRC32;

Query Match      58.1%; Score 61; DB 1; Length 1612;
Best Local Similarity 72.7%; Pred. No. 5,44e-01;
Matches      8:  Conservative      2:  Mismatches      1:  Indels      0:  Gaps      0:
Db           922 IFVVDNRIVEI 932
OY           3 IFVVDKNTGDI 13

RESULT 9
ID      TP2B-MOUSE      STANDARD:      PRT: 1612 AA.
AC      064511:
DT      15-JUL-1998 (REL. 36, CREATED)
DI      15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE      15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DN      DNA TOPOISOMERASE II, BETA ISOZYME (EC 5.99.1.3).
GN      TOP2B.
OS      MUS MUSCULUS (MOUSE).
OC      EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC      RODENTIA; SCIROGNATHI; MURIDAE; MURINAE; MUS.
RN      (1)
RP      SEQUENCE FROM N.A.
RC      STRAIN=BALE/C; TISSUE=BRAIN;
RL      MIYAIKE M., ADACHI N., KIKUCHI A.;
RL      SUBMITTED (SEP-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC      -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC      BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC      MAKES DOUBLE-STRAND BREAKS.
CC      -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC      OF DOUBLE-STRANDED DNA.
CC      -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
CC      -1- EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH NEGATIVE AND
CC      POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES RELAX ONLY
CC      NEGATIVE SUPERCOILS.
CC      -1- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: D38046; G1066004;
DR      MGD: M6196791; TOP2B.
DR      PROSITE: PS00177; TOPOISOMERASE-II; 1.
DR      PFAM: PF00204; DNA_topoisom; 1.
DR      HSSP: P06786; 1BGW.
KW      ISOMERASE; TOPOISOMERASE; DNA-BINDING; ATP-BINDING; NUCLEAR PROTEIN.
FT      ACT SITE 814 814 DNA CLEAVAGE (BY SIMILARITY).
SQ      SEQUENCE 1612 AA; 181863 MW; B496E4B5 CRC32;

Query Match      58.1%; Score 61; DB 1; Length 1612;
Best Local Similarity 72.7%; Pred. No. 5,44e-01;
Matches      8:  Conservative      2:  Mismatches      1:  Indels      0:  Gaps      0:
Db           922 IFVVDNRIVEI 932
OY           3 IFVVDKNTGDI 13

RESULT 10
ID      TP2B-HUMAN      STANDARD:      PRT: 1626 AA.
AC      Q02880:
DT      01-JUL-1993 (REL. 26, CREATED)
DI      01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DE      15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DN      DNA TOPOISOMERASE II, BETA ISOZYME (EC 5.99.1.3).
GN      TOP2B.
OS      HOMO SAPIENS (HUMAN).
OC      EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC      PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN      (1)
RP      SEQUENCE FROM N.A.
RC      MEDLINE: 93087165.
RX      JENKINS J.R., AYTON P., JONES T., DAVIES S.L., SIMMONS D.L.,
RX      HARRIS A.L., SHEER D., HICKSON I.D.;
RT      Isolation of cDNA clones encoding the beta isozyme of human DNA
RT      topoisomerase II and localisation of the gene to chromosome 3p24.
RL      NUCLEIC ACIDS RES. 20:5587-5592(1992).
RN      (2)
RP      SEQUENCE FROM N.A.
RX      MEDLINE: 93192319.
RA      AUSTIN C.A., SNG J.H., PATEL S., FISHER L.M.;
RT      "Novel Hela topoisomerase II is the II beta isoform: complete coding
RT      sequence and homology with other type II topoisomerases."
RL      BIOCIN. BIOPHYS. ACTA 1172:283-291(1993).
RN      (3)
RP      SEQUENCE OF 1038-1271 FROM N.A.

```

```

RX MEDLINE: 90306333.
RA AUSTIN C.A., FISHER L.M.;
RT "Isolation and characterization of a human cDNA clone encoding a
RL novel DNA topoisomerase II homologue from Hela cells.";
RL FEBS LETT. 266:115-117(1990).
RN [4]
RP ALTERNATIVE SPLICING.
RX MEDLINE: 93376494.
RA DAVIES S.L., JENKINS J.R., HICKSON I.D.;
RT "Human cells express two differentially spliced forms of
RL topoisomerase II beta mRNA.";
RL NUCLEIC ACIDS RES. 21:3719-3723(1993).
CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS.
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC OF DOUBLE-STRANDED DNA.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR, NUCLEOLUS.
CC -1- PTM: PHOSPHORYLATED.
CC -1- ENKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH NEGATIVE AND
CC POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES RELAX ONLY
CC NEGATIVE SUPERCOILS.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, CALLED BETA-1 AND BETA-2 ARE
CC PRODUCED BY ALTERNATIVE SPLICING OF THE TOP2B GENE. THE BETA-2
CC FORM IS SHOWN HERE.
CC -1- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X68060; G37231;
DR EMBL: X71911; -; NOT_ANNOTATED_CDS.
DR EMBL: Z15111; G288567;
DR EMBL: Z15115; G288565;
DR EMBL: X53662; G38325;
DR PIR: S26730; S26730.
DR PIR: S10710; S10710.
DR PIR: S41641; S41641.
DR MIM: 126431;
DR PROSITE: PS00177; TOPOISOMERASE_II; 1.
DR PFAM: PF00204; DNA_topoisomII; 1.
DR HSP: P06786; 1BGW.
KW ISOMERASE; TOPOISOMERASE; DNA-BINDING; ATP-BINDING; PHOSPHORYLATION;
KW ALTERNATIVE SPLICING; NUCLEAR PROTEIN.
FT ACT_SITE 826 826 DNA_CLEAVAGE (BY SIMILARITY).
FT VARSPIC 24 28 MISSING (IN BETA-1).
FT CONFLICT 1611 1611 T->A (IN REF. 2).
SQ SEQUENCE 1626 AA; 183296 MW; 4191FFD CRC32;
Query Match 58.1%; Score 61; DB 1; Length 1626;
Best Local Similarity 72.7%; Pred. No. 5.44e-01;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RA NIMI A., HARATA M., MIZUNO S.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC OF DOUBLE-STRANDED DNA.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR; GENERALLY LOCATED IN THE
CC NUCLEOLUS.
CC -1- ENKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH NEGATIVE AND
CC POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES RELAX ONLY
CC NEGATIVE SUPERCOILS.
CC -1- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB007446; D1023408;
DR PROSITE: PS00177; TOPOISOMERASE_II; 1.
DR PFAM: PF00204; DNA_topoisomII; 1.
DR HSP: P06786; 1BGW.
KW ISOMERASE; TOPOISOMERASE; DNA-BINDING; ATP-BINDING; NUCLEAR PROTEIN.
FT ACT_SITE 831 831 DNA_CLEAVAGE (BY SIMILARITY).
FT DOMAIN 5 23 POLY-GLY.
FT DOMAIN 1265 1268 POLY-GLY.
FT DOMAIN 1388 1391 POLY-ASP.
FT DOMAIN 1393 1396 POLY-ASN.
SQ SEQUENCE 1627 AA; 183245 MW; 4EEEAAB CRC32;
Query Match 58.1%; Score 61; DB 1; Length 1627;
Best Local Similarity 72.7%; Pred. No. 5.44e-01;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 934 IFVVDKNTVEI 944
| | | | | : | |
| | | | | : | |
QY 3 IFVVDKNTGDI 13

```

```

Db 939 IFVVDKNTVEI 949
| | | | | : | |
| | | | | : | |
QY 3 IFVVDKNTGDI 13

```

```

RESULT 11
ID TP2B_CHICK STANDARD: PRT; 1627 AA.
AC 042131;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE DNA TOPOISOMERASE II, BETA ISOZYME (EC 5.99.1.3).
GN TOP2B.
OS GALLUS GALLUS (CHICKEN).

```

```

RESULT 12
ID YRHB_ECOLI STANDARD: PRT; 94 AA.
AC P46857;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 10.6 KD PROTEIN IN GNT-R-GGT INTERGENIC REGION (094).
GN YRHB.
OS ESCHERICHIA COLI.
OS BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OS ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK B.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
SCIENCE 277:1453-1474(1997).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```
CC      modified and this statement is not removed. Usage by and for commercial  
CC      entities requires a license agreement (see http://www.isd-sib.ch/announce/  
CC      or send an email to license@isd-sib.ch).  
-----  
DR      EMBL; U18997; G606381; -.  
DR      EMBL; AE000421; G1789855; .  
DR      ECOGENE; EG12948; YRHB.  
KW      HYPOTHETICAL PROTEIN.  
SQ      SEQUENCE   94 AA: 10613 MW: 585316D CRC32;  
  
Query Match               57.1%; Score 60; DB 1; Length 94;  
Best Local Similarity    45.5%; Pred. No. 8,60e-01;  
Matches     5; Conservative    6; Mismatches    0; Indels    0; Gaps    0;  
  
Db       59 FIIKDSCGEIH 69  
         |:|::||:  
Oy        4 FYVDKNTGDIN 14  
  
RESULT   13  
ID CADE_HUMAN          STANDARD;             PRI;           790 AA.  
AC Q13634;  
DT 15-JUL-1998 (REL. 36, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE CADERHERIN-14 PRECURSOR.  
GN CDH14.  
OS Homo sapiens (human).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRhini; HOMINIDEA; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA SHIBATA T., SHIMOMURA Y., GOTOH M., HIROHASHI S.;  
RC TISSUE=Brain;  
RL SUBMITTED (JUN-1996) TO EMBL/GenBank/DDbj Data Banks.  
CL - FUNCTION: CADHERINS ARE CALCINIUM DEPENDENT CELL ADHESSION PROTEINS.  
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
CC SORTING OF HETEROGENEOUS CELL TYPES.  
CC - LOCATION: SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC - SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isd-sib.ch/announce/  
CC or send an email to license@isd-sib.ch).  
-----  
DR      EMBL; U59325; G1389853; .  
DR      PROSITE; PS00232; CAdherin_3.  
DR      Pfam; PF00028; cadherin_3.  
DR      Pfam; PF01049; Cadherin_C-term; 1.  
KW      HSSE; P15116; INCI.  
KW      CELL Adhesion; Glycoprotein; Phosphorylation; Transmembrane;  
KW      Calcium-binding; Repeat; Signal.  
FT      SIGNAL              1..24 POTENTIAL.  
FT      PROPEP              25..53 POTENTIAL.  
FT      CHAIN                54..790 CAdHerIn-14,  
FT      DOMAIN                609..636 EXTRACELLULAR (POTENTIAL),  
FT      TRANSMEM            637..790 POTENTIAL,  
FT      REPEAT                 54..159 CYTOPLASMIC (POTENTIAL),  
FT      REPEAT                 160..268 CAdHerIn 1,  
FT      REPEAT                 269..383 CAdHerIn 2,  
FT      REPEAT                 384..486 CAdHerIn 3,  
FT      REPEAT                 487..608 CAdHerIn 4,  
FT      CARBOHYD              36..36 CAdHerIn 5,  
FT      CARBOHYD             255..255 POTENTIAL,  
FT      CARBOHYD             455..455 POTENTIAL,  
FT      CARBOHYD             536..536 POTENTIAL.
```

```

SQ      SEQUENCE 790 AA; 88072 MW; 4596258F CRC32;
Query Match          57.1%; Score 60; DB 1; Length 790;
Best Local Similarity 58.3%; Pred. No. 8,60e-01;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db      100 IFIIDTGDH 111
      111:1 1111:
QY      3 IFVVDKNTGDI 14

RESULT 14
ID      POLG_P1YFV1 STANDARD; PRT; 3027 AA.
AC      Q05057;
DT      01-FEB-1994 (REL. 28, CREATED)
DT      01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT      15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE      GENOME POLYPROTEIN [CONTAINS: 22.5 KD PROTEIN; 26 KD PROTEIN; 31 KD
OS      PARAINFLUENZA VIRUS (ISOLATE P-121) (PYFV).
OC      VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; SEQUIVIRIDAE;
CC      SEQUIVIRUS.
CZ      [1]
RN      SEQUENCE FROM N.A.
RP      MEDLINE; 9310785.
RX      TURNBULDE-CROSS A.D., REAVY B., MAYO M.A., MURANT A.F.;
RA      "The nucleotide sequence of parainfluenza virus: a plant
RT      picorna-like virus.";
RL      J. GEN. VIROL. 73:3203-3211(1992).
CC      -I- SIMILARITY: SOME, TO THE CMV AND TERN POLYPROTEINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; D14066; G222457; -.
DR      PIR; J01917; J01917.
DR      PFM; PF00680; RNA_dep_RNA_pol; 1.
DR      PFM; PF00910; RNA_helicase; 1.
DR      KX      POLYPROTEIN; ATP-BINDING; COAT PROTEIN; TRANSFERASE;
KW      RNA-DIRECTED RNA POLYMERASE.
FT      NP_BIND 1467 1474 ATP (POTENTIAL).
FT      VARIANT 962 962 T -> I.
FT      VARIANT 1373 1373 L -> F.
SQ      SEQUENCE 3027 AA; 336242 MW; 18E7E4A CRC32;

Query Match          55.2%; Score 58; DB 1; Length 3027;
Best Local Similarity 40.0%; Pred. No. 2.12e+00;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db      2246 FGEVGNDTGDI 2260
      111:1 111:1
QY      1 FGEVGNDTGDI 15

RESULT 15
ID      RNH2_HELPY STANDARD; PRT; 209 AA.
AC      P56121;
DT      01-NOV-1997 (REL. 35, CREATED)
DT      01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT      01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE      RIBONUCLEASE HII (EC 3.1.26.4) (RNASE HII).
GN      RNH2 OR HP1323.
OS      HELICOBACTER PYLORI (CAMPILOBACTER PYLORI).
OC      BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; HELICOBACTER GROUP;
CC      HELICOBACTER.
CZ      [1]
RN      SEQUENCE FROM N.A.
RP      STRAIN-26695 / ATCC 700392;

```

```

RX MEDLINE: 97394467.
RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA NELSON K., OUDACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA LOEFVUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA VENTER J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL NATURE 388:539-547(1997).
CC -1- FUNCTION: DEGRADES THE RIBONUCLEOTIDE MOIETY ON RNA-DNA HYBRID
CC MOLECULES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ENDONUCLEOLYTIC CLEAVAGE TO 5'-PHOSPHO-
CC MONOESTER.
CC -1- SIMILARITY: WITH OTHER RNASE H FROM BACTERIA, FUNGI, AND WITH
CC RNASE H DOMAINS FROM POL. OF RETROVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE000634; G2314491;
DR TIGR: HP1323;
KW HYDROLASE; NUCLEASE; ENDONUCLEASE; MAGNESIUM
SQ SEQUENCE 209 AA; 23176 MW; FA2B072C CRC32;

```

```

Query Match          54.3%; Score 57; DB 1; Length 209;
Best Local Similarity 46.2%; Pred. No. 3.30e+00;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```

```

DB 67 GFFVYKKSANEID 79
OY 2 GFFVYKKSANEID 14

```

Search completed: Fri Jun 11 17:16:12 1999
Job time : 8 secs.

SEQUENCE FROM N.A.
 SIKAIN-WHITE LEGHORN; TISSUE-BRAIN;
 MEDLINE: 95309115.
 RA NAKAGAWA S., TAKEICHI M.;
 PT "Neural crest cell-cell adhesion controlled by sequential and
 subpopulation-specific expression of novel cadherins.";
 RL DEVELOPMENT 121:1321-1332(1995).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: D42150; D1008306; -
 DR PROSITE: PS00232; CADHERIN; 3.
 DR PFAM: PF00028; cadherin; 5.
 DR PFAM: PF01049; cadherin; C term; 1.
 KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
 SO SEQUENCE 785 AA; 87171 MW; F27A6881 CRC32;

Query Match 62.9%; Score 66; DB 13; Length 785;
 Best Local Similarity 61.5%; Pred. No. 1.08e-01;
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 93 SIFIDENGTGDIH 105
 :||:|||||:
 Qy 2 GIFVVDKNTGDI 14

RESULT 3
 ID 076356 PRELIMINARY; PRT: 1329 AA.
 AC 076356;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE C4567.6 PROTEIN.
 GN C4567.6.
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE: 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BOWFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES N., KERSHAW J., KIRSTEN J., LAISTER N., LATRILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SANDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., STURSTON J.,
 RA THERREY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL NATURE 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA DANTE M., WANSLEY P.;
 RT "The sequence of C. elegans cosmid C4567.";
 RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: AF067611; G3193145; -
 DR PROSITE: PS00232; CADHERIN; 1.
 KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
 SO SEQUENCE 1329 AA; 146518 MW; B161D39E CRC32;

Query Match 60.0%; Score 63; DB 5; Length 1329;
 Best Local Similarity 72.7%; Pred. No. 4.40e-01;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1288 IFVVDKSGDI 1298
 |||||:
 Qy 3 IFVVDKNTGDI 13

RESULT 4
 ID 024292 PRELIMINARY; PRT: 3380 AA.
 AC 024292;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE ADHERIN.
 GN DACHSOUS.
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EHYDROIDEA;
 OC DROSOPHILIDAE; DROSOPHILA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 95324813.
 RA CLARK H.F., BRENTNUP D., SCHNEITZ K., BIEBER A., GOODMAN C., NOLL M.;
 RT "Dachous encodes a member of the cadherin superfamily that controls
 imaginal disc morphogenesis in Drosophila.";
 RL GENES DEV. 9:1530-1542(1995).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: L08811; G685170; -
 DR FLYBASE: FBgn0000497; ds.
 DR PROSITE: PS00232; CADHERIN; 18.
 DR PFAM: PF00028; cadherin; 26.
 KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
 SO SEQUENCE 3380 AA; 366356 MW; 7D83FAC3 CRC32;

Query Match 58.1%; Score 61; DB 5; Length 3380;
 Best Local Similarity 58.3%; Pred. No. 1.10e-00;
 Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 713 GIFRIDSTGEI 724
 |||:|:|:
 Qy 2 GIFVVDKNTGDI 13

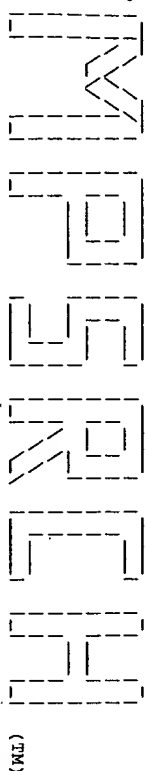
RESULT 5
 ID 055701 PRELIMINARY; PRT: 214 AA.
 AC 055701;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 24.5 KD PROTEIN.
 OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
 OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC6803;
 RA TABATA S.;
 RL SUBMITTED (AUG-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC6803;
 RX MEDLINE: 96127529.
 RA KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,
 RA SUGIURA M., TABATA S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 Synechocystis sp. strain PCC6803. I. sequence features in the 1Mb
 RT region from map positions 64% to 92% of the genome.";
 RL DNA RES. 2:153-166(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC6803;
 RX MEDLINE: 97061201.
 RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
 RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAKOTO S., KIMURA T.,
 RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARO K., OKUMURA S.,
 RA SHIMO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M.,

RA TABATA S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. PCC6803. II. Sequence determination of the entire
 RT genome and assignment of potential protein-coding regions.";
 RL DNA RES. 3:109-136(1996).
 DR EMBL; D64000; D1010882;
 KM HYPOTHETICAL PROTEIN.
 SO SEQUENCE 214 AA; 24539 MW; CDEA936D CRC32;
 DB 49 FGIVMDPSTGEIS 62
 OY 1 FGIVFVDMKNTGDIN 14
 Query Match 57.1%; Score 60; DB 2; Length 214;
 Best Local Similarity 42.9%; Pred. No. 1.72e+00;
 Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
 RESULT 6 PRELIMINARY; PRT; 598 AA.
 AC 029807;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE ALDEHYDE PEROXIDASE OXIDOREDUCTASE (AOR-3).
 GN AF0340.
 OS ARCHAEOGLOBUS FULGIDUS.
 OC ARCHAEA; EURYARCHAEOTA; ARCHAEOGLOBALES; ARCHAEOGLOBACEAE;
 CC ARCHAEOGLOBUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE; 98049343.
 KA KLEINK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
 KA KUTCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
 KA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
 KA FLEISCHMANN R.D., DUCKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
 KA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
 KA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODER A., ZHOU L.,
 KA OVERBEER R., GOCAYNE J.D., WEIDMAN J.F., MCDONALD L., UTTERBACK T.,
 KA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
 KA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
 KA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
 KA VENTER J.C.;
 KI "The complete genome sequence of the hyperthermophilic,
 KI sulphate-reducing archaeon Archaeoglobus fulgidus.";
 RT NATURE 390:364-370(1997).
 RL EMBL; AE001081; G2650295;
 DR TIGR; AF0340;
 DR PFAM; PF01314; AFOR; 1.
 KM HYPOTHETICAL PROTEIN.
 SO SEQUENCE 598 AA; 65665 MW; 909BAB0E CRC32;
 DB 355 FAMYLTENGVDENI 369
 OY 1 FGIVFVDMKNTGDINI 15
 Query Match 57.1%; Score 60; DB 1; Length 598;
 Best Local Similarity 40.0%; Pred. No. 1.72e+00;
 Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;
 RESULT 7 PRELIMINARY; PRT; 154 AA.
 AC 058735;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL PROTEIN MJ1339.
 GN MJ1339.
 OS METHANOCOCCUS JANNASCHII.
 OC ARCHAEA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE;
 CC METHANOCOCCUS.

RN [1]
 RX SEQUENCE FROM N.A.
 RX MEDLINE; 96337999.
 RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
 RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
 RA OVERBEER R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
 RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
 RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
 RA KLEINK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschii.";
 RL SCIENCE 273:1058-1073(1996).
 DR EMBL; U67574; G1591981;
 KM HYPOTHETICAL PROTEIN.
 SO SEQUENCE 154 AA; 17196 MW; 05503C2B CRC32;
 DB 106 YGIFIKTVGDID 119
 OY 1 FGIVFVDMKNTGDIN 14
 Query Match 56.2%; Score 59; DB 1; Length 154;
 Best Local Similarity 42.9%; Pred. No. 2.68e+00;
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 RESULT 8 PRELIMINARY; PRT; 245 AA.
 AC 058723;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL PROTEIN MJ1327.
 GN MJ1327.
 OS METHANOCOCCUS JANNASCHII.
 OC ARCHAEA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE;
 CC METHANOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE; 96337999.
 RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
 RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
 RA OVERBEER R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
 RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
 RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
 RA KLEINK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschii.";
 RL SCIENCE 273:1058-1073(1996).
 DR EMBL; U67573; G1591968;
 KM HYPOTHETICAL PROTEIN.
 SO SEQUENCE 245 AA; 28779 MW; 3D81831F CRC32;
 DB 151 YGIFVVDK 158
 OY 1 FGIVFVDK 8
 Query Match 56.2%; Score 59; DB 1; Length 245;
 Best Local Similarity 87.5%; Pred. No. 2.68e+00;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 RESULT 9 PRELIMINARY; PRT; 624 AA.
 AC 048662;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE PV72.
 OS CUCURBITA MAXIMA (PUMPKIN) (WINTER SQUASH).

OC EUKARYOTA: VIRIDIPLANTAE; STREPTOPHYTA; EMEROPHYTA; TRACHEOPHYTA;
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC CUCURBITALES; CUCURBITACEAE; CUCURBITA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-DEVELOPING COTYLEDON;
 RX MEDLINE; 98182943.
 RA SHIMADA T., KUROYANAGI M., NISHIMURA M., HARA-NISHIMURA I.;
 RT "A pumpkin 72-kDa membrane protein of precursor-accumulating vesicles
 has characteristics of a vacuolar sorting receptor.";
 RL PLANT CELL PHYSIOL. 38:1414-1420(1997).
 DR EMBL; AB006809; D1026005; -
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 KM GLYCOPROTEIN; EGF-LIKE DOMAIN.
 SQ SEQUENCE 624 AA; 69028 MW; AAA97DOB CRC32;
 Query Match 56.2%; Score 59; DB 10; Length 624;
 Best Local Similarity 70.0%; Pred. No. 2.68e+00;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 20 FGFVVDKNT 10
 QY 1 FGFVVDKNT 10
 RESULT 10
 ID 033373 PRELIMINARY; PRT; 1033 AA.
 AC 033373;
 DT 01-JAN-1998 (TREMELREL. 05, CREATED)
 DT 01-JAN-1998 (TREMELREL. 05, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMELREL. 07, LAST ANNOTATION UPDATE)
 DE PILC PROTEIN.
 GN PILC.
 OS NEISSERIA GONORRHOEA.
 OC BACTERIA; PROTEOBACTERIA; BETA SUBDIVISION; NEISSERIACEAE; NEISSERIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-GC-653;
 RA BAECKMAN M., KALLSTROM H., JONSSON A.B.;
 RL MICROBIOLOGY 144:149-156(1998).
 DR EMBL; AJ001121; E334755; -
 SQ SEQUENCE 1033 AA; 112936 MW; B0593053 CRC32;
 Query Match 56.2%; Score 59; DB 2; Length 1033;
 Best Local Similarity 46.7%; Pred. No. 2.68e+00;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 Db 804 YGIFDDKGTGVK 818
 QY 1 FGFVVDKNTGIDINI 15
 RESULT 11
 ID 028514 PRELIMINARY; PRT; 270 AA.
 AC 028514;
 DT 01-JAN-1998 (TREMELREL. 05, CREATED)
 DT 01-JAN-1998 (TREMELREL. 05, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMELREL. 07, LAST ANNOTATION UPDATE)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN AF1760.
 OS ARCHAEoglobus fulgidus.
 OC ARCHAEA; EURARCHAEOTA; ARCHAEoglobales; ARCHAEoglobaceae;
 CC ARCHAEoglobus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE; 98049343.
 RA KLEINK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
 RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
 RA KERCHOW D.L., KERLAVAGE A.R., GRAHAM D.E., KYRIDES N.C.,
 RA FLEISCHMAN R.D., OJACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
 RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,

RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
 RA OVERBERK R., GOCAINE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
 RA COTTON M.D., SPRIGGS T., ARTACH P., KAINE B.P., SYKES S.M.,
 RA SADOW P.M., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
 RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., MOESE C.R.,
 RA VENTER J.C.;
 RT "The complete genome sequence of the hyperthermophilic,
 RT sulphate-reducing archaeon Archaeoglobus fulgidus.";
 RL NATURE 390:364-370(1997).
 DR EMBL; AE000981; G2648791; -
 DR TIGR; AF1760; -
 KM HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 270 AA; 29554 MW; E5F8A9B CRC32;
 Query Match 55.2%; Score 58; DB 1; Length 270;
 Best Local Similarity 40.0%; Pred. No. 4.15e+00;
 Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
 Db 165 YAKTVVKTGGEIHI 179
 QY 1 FGFVVDKNTGIDINI 15
 RESULT 12
 ID 065522 PRELIMINARY; PRT; 593 AA.
 AC 065522;
 DT 01-AUG-1998 (TREMELREL. 07, CREATED)
 DT 01-AUG-1998 (TREMELREL. 07, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMELREL. 07, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 68.6 KD PROTEIN.
 GN FAD11.10.
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
 OC EUKARYOTA: VIRIDIPLANTAE; STREPTOPHYTA; EMEROPHYTA; TRACHEOPHYTA;
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BEVAN M., BENES V., RECHMANN S., BORKOVA D., ANSORGE W., HOHEISEL J.,
 RA MEWES H.W., MAYER K.F.X., SCHUELLER C.;
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU ARABIDOPSIS SEQUENCING PROJECT;
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AL022537; E1287270; -
 KM HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 593 AA; 68632 MW; 9EA10B88 CRC32;
 Query Match 55.2%; Score 58; DB 10; Length 593;
 Best Local Similarity 35.7%; Pred. No. 4.15e+00;
 Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;
 Db 528 FAVFLEKIDIPDLK 541
 QY 1 FGFVVDKNTGIDINI 14
 RESULT 13
 ID 075284 PRELIMINARY; PRT; 816 AA.
 AC 075284;
 DT 01-NOV-1998 (TREMELREL. 08, CREATED)
 DT 01-NOV-1998 (TREMELREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)
 DE KIA0345-Like 9.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
 CC CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA KIMMERLY W., BONDUG M., CHENG J., CONNOLLY K.S., GUNNING K.M.,
 RA KAMNER K., MGUEL T., MILLER C., PILLUCK S., POLLARD M., ROJESKI H.,
 RA SUBRAMANIAN S., MARTIN C.H.;
 RT "Sequencing of human chromosome 5.";



Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 11 17:27:18 1999; Maspar time 4.80 Seconds
66.422 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-3
Description: (1-15) from US08991628.pep
Perfect Score: 96
Sequence: 1 LNSKIAKIVSQEPA 15
Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 18.085; Variance 51.973; scale 0.348

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	96	100.0	15 35	W78814	Desmoglein 3 protein	3.65e-04
2	96	100.0	15 20	W04843	Self epitope of desmo	3.65e-04
3	96	100.0	15 33	W64815	Desmoglein-3 190-204.	3.65e-04
4	96	100.0	614 19	W07908	Pemphigus vulgaris an	3.65e-04
5	96	100.0	999 6	R30742	Human pemphigus vulga	3.65e-04
6	87	90.6	778 21	W15489	Pemphigus foliaceus a	5.06e-03
7	77	80.2	263 24	W13010	Segment of desmosomal	8.83e-02
8	77	80.2	560 24	W13009	Segment of desmosomal	8.83e-02
9	54	56.3	878 10	R53060	Sequence of human liv	4.45e+01
10	54	56.3	878 15	R85487	Human E-cadherin prec	4.45e+01
11	54	56.3	2233 33	W48711	HPV-3 JS isolate W11	4.45e+01
12	54	56.3	2233 33	W48712	HPV-3 FRH1 CP45 vacc	4.45e+01
13	54	56.3	2233 33	W48713	HPV-3 Vero CP45 vacc	4.45e+01
14	53	55.2	456 37	W82841	Human cerebral protei	5.74e+01
15	49	51.0	572 8	W40843	Bilirubin oxidase.	1.57e+02
16	49	51.0	916 21	W13129	Full length human cad	1.57e+02

17	49	51.0	916 24	W25658	Human cadherin-4.	1.57e+02
18	49	51.0	2408 2	R24307	Translation of ORF 3	1.57e+02
19	48	50.0	26 32	W60192	Bacteriophage spo1 Po	2.01e+02
20	48	50.0	528 22	W20908	H. pylori inner membr	2.01e+02
21	48	50.0	2016 30	W23994	Human hH1 sodium chan	2.01e+02
22	48	50.0	2019 13	R67913	Cardiac sodium channe	2.01e+02
23	48	50.0	2020 2	R06584	Cardiac sodium channe	2.01e+02
24	47	49.0	198 1	R06643	Mammalian growth hormo	2.56e+02
25	47	49.0	308 36	W75006	Human secreted protei	2.56e+02
26	47	49.0	330 39	W89745	Staphylococcus aureus	2.56e+02
27	47	49.0	339 37	W74883	Human secreted protei	2.56e+02
28	47	49.0	573 17	R75741	B11 outer surface pro	2.56e+02
29	47	49.0	700 17	R75731	B. burgdorferi strain	2.56e+02
30	47	49.0	726 6	R30730	B. burgdorferi 79 KD	2.56e+02
31	47	49.0	756 6	R30177	P.T.	2.56e+02
32	47	49.0	4472 19	R97246	Yarrowence gene cluste	2.56e+02
33	46	47.9	33 35	W79334	Staphylococcus aureus	3.26e+02
34	46	47.9	298 18	R97629	Human SLAMF2 T-cell co	3.26e+02
35	46	47.9	305 18	R97630	Human SLAMF3 T-cell co	3.26e+02
36	46	47.9	307 18	R97631	Human SLAMF4 T-cell co	3.26e+02
37	46	47.9	335 18	R97628	Human SLAMF1 T-cell co	3.26e+02
38	46	47.9	481 26	W34554	MITL glycosidase 29G	3.26e+02
39	46	47.9	481 36	W49858	Desulfococcus MITL	3.26e+02
40	46	47.9	486 29	W38455	Yeast RNA-binding pro	3.26e+02
41	46	47.9	779 18	R98326	Rat neuronal protein	3.26e+02
42	46	47.9	821 19	R99579	Calpain large subunit	3.26e+02
43	46	47.9	1275 5	R13516	P. dentrificans COB N	3.26e+02
44	46	47.9	1276 7	R35199	Mouse multigene resis	3.26e+02
45	46	47.9	2440 22	W20828	H. pylori cytoplasmic	3.26e+02

ALIGNMENTS

RESULT 1
ID W78814: standard; peptide; 15 AA.
AC W78814:
DT 17-NOV-1998 (first entry)
DE Desmoglein 3 protein fragment 190-204.
KW Microparticle; delivery; polymeric matrix; autoantigen; tumour antigen;
class II associated peptide; pathogen; gene therapy; genetic disease;
infection; downregulation; immune response.
OS Homo sapiens.
PN Synthetic.
PS W09831398-A1.
PD 23-JUL-1998.
PF 22-JAN-1998; U01499.
PI 06-JAN-1997; US-003253.
PR 22-JAN-1997; US-787547.
PA (PANG-) PANGAEN PHARM INC.
PI Curley JM, Hedley ML, Langer RS, Lunsford LB;
WPI; 98-427556/36.
DR New preparations of microparticles - comprising a synthetic polymer
matrix and nucleic acid comprising an expression vector for use in
gene therapy
PT gene therapy
PS Disclosure; Page 8, 101pp; English.
CC A microparticle preparation (MP) has been developed, consisting of
microparticles having a diameter of less than 100 nm. The MP comprises:
(a) a polymeric matrix (PM) consisting of one or more synthetic polymers
having a solubility in water of less than 1 mg/l; and (b) an expression
vector selected from RNA molecules (at least 50% of which are supercoiled).
CC Also described is a MP of at most 20 microns in diameter, comprising: (a)
a PM; and (b) a NM comprising an expression control sequence operatively
linked to a coding sequence, where the coding sequence encodes an
expression product selected from: (1) a polypeptide at least 7 amino
acids in length, having a sequence identical to the sequence of: (i) a
fragment of a naturally-occurring mammalian protein; or (ii) a fragment
of a naturally-occurring protein from an infectious agent which infects
a mammal; (2) a peptide having a length and sequence which permits it to
bind to an MHC class I or II molecule; and (3) the polypeptide or the
peptide linked to a trafficking sequence. W69763 to W69765, and W78793
CC to W78897 are peptide fragments for use in the present invention. The
MPS are highly effective vehicles for the delivery of polynucleotides

CC into phagocytic cells. They can be used for gene therapy, e.g. for
 CC treating genetic diseases, infections or tumours or for downregulating
 CC an immune response.
 SQ Sequence 15 AA;

Query Match 100.0%; Score 96; DB 35; Length 15;
 Best Local Similarity 100.0%; Pred. No. 3.65e-04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 lnskiafkivsgpea 15
 |||||||
 QY 1 LNSKIAFKIVSOEPA 15

RESULT 2
 ID W04843 standard; peptide; 15 AA.
 AC W04843:

DE 18-FEB-1997 (first entry)
 KW Self epitope of desmoglein 3, implicated in autoimmune disease.
 KW Tolerantation; self-epitope; antigen; autoimmune disease;
 KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
 KW pemphigus vulgaris; desmoglein; multiple sclerosis;
 KW herpes simplex virus; adenovirus; phosphomannomutase;
 KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
 KW Influenza; haemagglutinin; reovirus; sigma protein.
 OS Homo sapiens.
 PN W09627387-A1.
 PD 12-SEP-1996.
 PR 07-MAR-1996; U03182.
 PA (HARD) HARVARD COLLEGE.
 PI Strominger JL, Mucherfennig KW;
 DR WPI: 96-425218/42.

PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
 PT antigens - useful in disease treatment, and method for
 PT identification of other self and non-self antigens implicated in
 PT auto-immune disease
 PS Claim 1; Page 39; 58pp; English.
 CC Pharmaceutical preparations for toleration to antigens comprise
 CC either an isolated human non-collagen or non-mysin basic protein
 CC (MBP) polypeptide which is capable of tolerising an individual to an
 CC autoantigen; or an isolated human pathogen polypeptide capable of
 CC tolerising an individual to that polypeptide. In both cases, the
 CC polypeptide (whether self or non-self) includes an amino acid
 CC sequence corresponding to a sequence motif for a MHC class II
 CC protein, such as HLA-DR, which is associated with a human autoimmune
 CC disease and which binds to the polypeptide to activate autoreactive
 CC T-cells in individuals with the autoimmune disease. This peptide is
 CC derived from the human desmoglein 3 protein (amino acids 190-204)
 CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
 CC derived from the human desmoglein protein are described in W04841-47.
 SQ Sequence 15 AA;

Query Match 100.0%; Score 96; DB 20; Length 15;
 Best Local Similarity 100.0%; Pred. No. 3.65e-04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 lnskiafkivsgpea 15
 |||||||
 QY 1 LNSKIAFKIVSOEPA 15

RESULT 3
 ID W64815 standard; peptide; 15 AA.
 AC W64815:

DE 29-SEP-1998 (first entry)
 KW Desmoglein-3 190-204.
 KW autoantigen; Dg; gene therapy; pemphigus vulgaris; microparticle;
 OS Homo sapiens.
 PN US5783567-A.
 PD 21-JUL-1998.
 PR 22-JAN-1997; 787547.

PR 22-JAN-1997; US-787547.
 PA (PANG-) PANGAEA PHARM INC.
 PI Curley JM, Hedley ML, Langer RS;
 DR WPI: 98-427077/36.

PT Microparticle encapsulated nucleic acids - for recombinant
 PT expression of proteins e.g. in gene therapy
 PS Disclosure; Column 4; 42pp; English.
 CC The patent describes a new preparation of microparticles each
 CC comprising a polymeric matrix and a nucleic acid. The polymeric
 CC matrix consists of one or more synthetic polymers having a solubility
 CC in water of less than 1 mg/l (e.g. poly-lactic-co-glycolic acid);
 CC and at least 90% of the microparticles have a diameter of less than
 CC 100 microns. The microparticles are useful for the delivery of nucleic
 CC acids to phagocytic cells. In one embodiment the microparticles are
 CC less than 20 microns in diameter and the nucleic acid (preferably in
 CC closed circular form) includes an expression control sequence
 CC operatively linked to a coding sequence, where the expression product
 CC of the coding sequence is a polypeptide having a length and a sequence
 CC which permits it to bind to an MHC class I or II molecule. The
 CC expression product is thus an effective stimulator of an immune
 CC response in mammals. The present sequence, an antigenic portion of
 CC desmoglein 3, is an example of an MHC class II peptide which can be
 CC expressed by the nucleic acid. It is associated with pemphigus
 CC vulgaris.
 SQ Sequence 15 AA;

Query Match 100.0%; Score 96; DB 33; Length 15;
 Best Local Similarity 100.0%; Pred. No. 3.65e-04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 lnskiafkivsgpea 15
 |||||||
 QY 1 LNSKIAFKIVSOEPA 15

RESULT 4
 ID W07908 standard; protein; 614 AA.
 AC W07908:

DE 29-JAN-1997 (first entry)
 KW Pemphigus vulgaris antigen protein extracellular region.
 KW Autoantibody; immunoglobulin G; IgG1; fusion protein; diagnosis;
 KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;
 KW dermatology.
 OS Homo sapiens.
 PN J08188540-A.
 PD 23-JUL-1996.
 PR 30-JUN-1995; 165632.
 PA (NISH/) NISHUKAWA T.
 DR WPI: 96-388562/39.
 PT Fused protein recognised by pemphigus vulgaris auto-antibody -
 PT useful to treat and diagnose pemphigus vulgaris
 PS Claim 1; Page 7-9; 9pp; Japanese.
 CC W07908 represents the human pemphigus vulgaris (PV) antigen
 CC extracellular region. The PV antigen is produced in patients with
 CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare
 CC relapsing disease causing suprabasal, intra-epidermal bullae
 CC (vesicles) of the skin and mucous membranes, which is fatal if
 CC untreated. The PV antigen was fused to a human IgG1 hinge region
 CC and the resulting fusion protein is useful to treat or diagnose
 CC pemphigus vulgaris.
 SQ Sequence 614 AA;

Query Match 100.0%; Score 96; DB 19; Length 614;
 Best Local Similarity 100.0%; Pred. No. 3.65e-04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 189 lnskiafkivsgpea 203
 |||||||
 QY 1 LNSKIAFKIVSOEPA 15

RESULT 5

ID R30742 standard; Protein: 999 AA.
AC R30742:
DE 14-JUN-1993 (first entry)
KW Human pemphigus vulgaris 130kD antigen.
KW Pemphigus vulgaris; skin disease; autoantibodies;
KW Keratinocyte cell surface antigen; glycoprotein; cell adhesion.
OS Homo sapiens.
PN US7798918-A.
PD 15-DEC-1992.
PF 27-NOV-1991; 798918.
PR (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Amagai M, Klaus-Kovtun V, Stanley JR;
WPI: 93-067436/08.
DR N-PSDB: Q35992.
RT DNA encoding pemphigus vulgaris antigen - useful in proteins for
RT diagnostic and therapeutic uses
PS Disclosure: Fig 7; 50pp; English.
CC This sequence is the pemphigus vulgaris 130kD antigen. The protein
CC and its encoding DNA may be used in the diagnosis and treatment of
CC pemphigus vulgaris. It is thought that the antigen may be a cell
CC adhesion molecule.
SQ Sequence 999 AA;

Query Match 100.0%; Score 96; DB 6; Length 999;
Best Local Similarity 100.0%; Pred. No. 3.65e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 190 lnskiafkivsgpea 204
OY 1 lnskiafkivsgpea 15
|||||
RESULT 6
ID W15489 standard; Protein: 778 AA.
AC W15489:
DE 17-JUN-1997 (first entry)
KW Pemphigus foliaceus antigen-IgG constant region fusion protein.
KW Pemphigus foliaceus; autoantibody; constant region; IgG;
KW extracellular region; antigen; hinge portion; skin;
KW dermatitis herpetiformis; fusion protein; detection; ss.
OS Chimeric - Homo sapiens.
FH Key location/Qualifiers
FT domain 1..545 "Pemphigus foliaceus antigen protein"
FT J09077800-A.
PN 25-MAR-1997.
PD 12-SEP-1995; 260899.
PF 12-SEP-1995; JP-260899.
PR (NISH/) NISHIKAWA T.
PA WPI: 97-241758/22.
DR P-PSDB: T66428.
RT Pemphigus foliaceus antigen-IgG constant region fusion protein - linked
RT through the hinge region used to treat pemphigus foliaceus
PS Claim 1; Page 10-12; 17pp; Japanese.
CC This sequence represents a fused protein recognised by pemphigus
CC foliaceus patient autoantibody which comprises the constant region
CC of IgG linked to the extracellular region of pemphigus foliaceus
CC antigen protein through the hinge portion. Pemphigus foliaceus is
CC a chronic, generalised, vesicular and scaling skin eruption similar
CC to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion
CC protein is useful to treat pemphigus foliaceus. The antigen is
CC especially administered through an adsorbent upon which the fusion
CC protein is immobilised via a carrier. The fusion protein is also
CC useful for detecting pemphigus foliaceus antibodies which is useful
CC in immunodiagnosis. The fusion protein has little or no side effects.
SQ Sequence 778 AA;

Query Match 90.6%; Score 87; DB 21; Length 778;
Best Local Similarity 80.0%; Pred. No. 5.06e-03;
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 190 lnskiafkivsgpea 204

OY 1 lnskiafkivsgpea 15
|||||
RESULT 7
ID W13010 standard; Protein: 263 AA.
AC W13010:
DE 21-NOV-1997 (first entry)
KW Segment of desmosomal cadherin, desmoglein Dsg2.
KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;
KW micrometastasis; separation; enrichment; targeted delivery;
KW metastatic.
OS Homo sapiens.
PN DE19531033-A1.
PD 27-FEB-1997.
PF 23-AUG-1995; 031033.
PR (PROG-) PROGEM BIOTECHNIK GMBH.
PI Franke WW, Schaefer S;
WPI: 97-146518/14.
DR Anti-body reactive with part of desmosomal cadherin - exposed on
RT surface of epithelial or carcinoma cells, not bound to desmosomes,
RT useful for diagnosis and treatment of carcinoma micrometastases
PS Claim 9; Page 5; 8pp; German.
CC The present sequence is a segment of the desmosomal cadherin (DC),
CC desmoglein Dsg2, which is exposed on the surface of epithelial or
CC carcinoma cells and not bound to desmosomes. An antibody (Ab)
CC directed against epitopes of the present sequence can be used to
CC diagnose, i.e. to detect carcinoma cells, especially
CC micrometastases, not bound to desmosomes, to separate, enrich or
CC detect living or fixed carcinoma cells by cell sorting methods and
CC as a therapeutic to deliver agents, e.g. other Ab or toxins, to
CC target cells. The Ab provides rapid and reliable detection of
CC metastatic carcinoma, and detects parts of DC that are not
CC accessible in desmosome bound cells, as in normal tissue or
CC carcinomas.
SQ Sequence 263 AA;

Query Match 80.2%; Score 77; DB 24; Length 263;
Best Local Similarity 73.3%; Pred. No. 8.83e-02;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 41 lnskisytyvslepa 55
OY 1 lnskiafkivsgpea 15
|||||
RESULT 8
ID W13009 standard; Protein: 560 AA.
AC W13009:
DE 21-NOV-1997 (first entry)
KW Segment of desmosomal cadherin, desmoglein Dsg2.
KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;
KW micrometastasis; separation; enrichment; targeted delivery;
KW metastatic.
OS Homo sapiens.
PN DE19531033-A1.
PD 27-FEB-1997.
PF 23-AUG-1995; 031033.
PR (PROG-) PROGEM BIOTECHNIK GMBH.
PI Franke WW, Schaefer S;
WPI: 97-146518/14.
DR Antibody reactive with part of desmosomal cadherin - exposed on
RT surface of epithelial or carcinoma cells, not bound to desmosomes,
RT useful for diagnosis and treatment of carcinoma micrometastases
PS Claim 7; Page 5; 8pp; German.
CC The present sequence is a segment of the desmosomal cadherin (DC),
CC desmoglein Dsg2, which is exposed on the surface of epithelial or
CC carcinoma cells and not bound to desmosomes. An antibody (Ab)
CC directed against epitopes of the present sequence can be used to

CC diagnose, i.e. to detect carcinoma cells, especially
 CC micrometastases, not bound to desmosomes, to separate, enrich or
 CC detect living or fixed carcinoma cells by cell sorting methods and
 CC as a therapeutic to deliver agents, e.g. other Ab or toxins, to
 CC target cells. The Ab provides rapid and reliable detection of
 CC metastatic carcinoma, and detects parts of DC that are not
 CC accessible in desmosome bound cells, as in normal tissue or
 CC carcinomas.
 CC Sequence 560 AA;

Query Match 80.2%; Score 77; DB 24; Length 560;
 Best Local Similarity 73.3%; Pred. NO. 8.83e-02;
 Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DB 143 Inskisytyvslepa 157
 1 LNSKIARFIVSOEP 15

RESULT 9
 ID R5060 standard; Protein; 878 AA.
 AC R5060;
 DT 08-NOV-1994 (first entry)
 DE Sequence of human liver E-cadherin.
 KW Human epithelial-cadherin; E-cadherin; cell adhesion molecule; CAM;
 KW uvomorulin; L-CAM; Cell CAM 120/80.
 OS Homo sapiens.
 PN MO9411401-A.
 PD 26-MAY-1994.
 PE 16-NOV-1993; U11097.
 PR 17-NOV-1992; US-978897.
 PA (UYXA) UNIV YALE.
 PI Morrow JS, Rimm DL;
 DR WPI; 94-183426/22.
 DR N-PSDB: 065487.
 PT Purified human E-cadherin protein and nucleic acid - used to
 PT develop prods. for diagnosis, prognosis, therapy and prophylaxis
 PT of E-cadherin disorders, e.g. malignancies
 PS Claim 1; Page 59-63; 97pp; English.
 CC E-cadherin is a cell adhesion molecule that is also known as
 CC uvomorulin, L-CAM and Cell CAM 120/80. The DNA encoding hEC was obt.
 CC by screening normal human liver and hepatocellular carcinoma cDNA
 CC libraries and a colonic epithelial cell cDNA library. The following
 CC sequences are specifically claimed: AAs 1-878; 151-878; 30
 CC sequential AAs from AAs 308-878; AAs 1-150; AAs 178-289; AAs 290-
 CC 401; AAs 402-513; AAs 178-513; AAs 151-703; AAs 1-703; AAs 728-878;
 CC AAs 704-878; nucleotide sequences comprising nucleotide numbers 116-
 CC 2749; 566-2749; 1-1053; 510-2696; 1332-3000; 540-1500; 348-906; 880-
 CC 1448; 384-1208; 641-2046; 685-1336; 880-1661; 1199-1742; 1373-1742;
 CC 1705-2204; 2458-2775; DNA encoding at least 30 AAs selected from
 CC AAs 308-878. The prods. can be used in the diagnosis, prognosis,
 CC therapy and prophylaxis of conditions involving improper E-cadherin
 CC expression. Suitable dosages for i.v. admin. of a protein are
 CC 20-500 mcg/kg body wt.
 CC Sequence 878 AA;
 SQ

Query Match

Best Local Similarity 53.3%; Score 54; DB 10; Length 878;

Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

DB 293 naalaytlisqdp 305
 2 NSKIARFIVSOEP 14

RESULT 10
 ID R85487 standard; Protein; 878 AA.
 AC R85487;
 DT 18-MAR-1996 (first entry)
 DE Human E-cadherin precursor.
 KW E-cadherin; T-lymphocyte; alpha-E,beta-7 integrin; cell adhesion;
 KW autoimmune disease; Crohn disease; psoriasis.
 OS Homo sapiens.

PH Key Location/Qualifiers
 FT peptide 1..150
 FT /label= Sig_peptide
 FT domain 151..702
 FT /label= Extracellular domain
 FT /note= "the extracellular domain (amino acids 1-552
 FT of the mature protein) is the preferred
 FT region for generation of peptides of the
 FT invention"

FT domain 703..726
 FT /label= Transmembrane_domain
 FT /label= Cytoplasmic_domain

PN WO9529693-A1.
 PD 09-NOV-1995.
 PE 03-MAY-1995; U05518.
 PR 03-MAY-1994; US-237919.
 PA (BGM) BRIGHAM & WOMEN'S HOSPITAL.
 PI Brenner MB, Cepex KL;
 DR WPI; 95-392921/50.
 DR N-PSDB: T05764.
 PT Inhibiting adhesion of T lymphocytes with E-cadherin - useful for
 PT isolating agents to treat auto-immune diseases e.g. Crohn's disease,
 PT psoriasis, etc

PS Disclosure; Page 70-75; 103pp; English.
 CC The human E-cadherin protein precursor (R85487) is expressed by
 CC an cDNA clone (T05764) derived from human liver. The extracellular
 CC domain of E-cadherin is used to generate peptides that specifically
 CC bind to heterocyclic cognates of E-cadherin and which inhibit adhesion
 CC of intra-epithelial T-lymphocytes to E-cadherin-expressing epithelial
 CC or endothelial cells in vitro or in vivo, thereby modulating
 CC mucosal immune responses. Such peptides are also specifically
 CC reactive with a monoclonal antibody (E4.6 or E6.1) that binds to
 CC E-cadherin and that can inhibit T-cell binding.
 CC Sequence 878 AA;
 SQ

Query Match 56.3%; Score 54; DB 15; Length 878;
 Best Local Similarity 53.8%; Pred. NO. 4.45e+01;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

DB 293 naalaytlisqdp 305
 2 NSKIARFIVSOEP 14

RESULT 11
 ID W48711 standard; Protein; 2233 AA.
 AC W48711;
 DT 13-OCT-1998 (first entry)
 DE HIV-3 US isolate wild-type L protein.
 KW L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
 KW single stranded RNA virus; Mononegavirales.
 OS Human parainfluenza virus.

PN WO9813501-A2.
 PD 02-APR-1998.
 PE 19-SEP-1997; U16718.
 PR 27-SEP-1996; US-026823.
 PA (AMCY) AMERICAN CYANAMID CO.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
 DR WPI; 98-230710/20.
 DR N-PSDB: V18272.
 PT Recombinantly-generated, attenuated, non-segmented, negative-sense,
 PT single stranded RNA virus of order Mononegavirales - having
 PT attenuating mutation in 3' genomic promoter region and RNA

PT polymerase gene, useful as vaccine to immunise against such virus
 PS Disclosure; Page 246-254; 426pp; English.
 CC This sequence represents the wild-type L protein from Human parainfluenza
 CC virus (HPIV-3) type 3 isolate US. This sequence is used in a method which
 CC involves the isolation of recombinantly-generated, attenuated,
 CC non-segmented, negative-sense, single stranded RNA virus of the order
 CC Mononegavirales which have at least 1 attenuating mutation in the 3'

CC genomic promoter region and at least 1 attenuating mutation in the RNA

```
CC polymerase gene. This RNA virus can be used as a vaccine to immunise an
CC individual against such a virus.
SQ Sequence 2233 AA;

Query Match          56.3%; Score 54; DB 33; Length 2233;
Best Local Similarity 33.3%; Pred. No. 4.45e+01;
Matches      5; Conservative       7; Mismatches    3; Indels     0; Gaps     0;

Db      973 ldrsvlyrimnqepg 987
      | : : : : | : ||| :
Oy      1 LNSKIAFRIVSOEPA 15

RESULT   12
ID W48712 standard; Protein; 2233 AA.
AC W48712;
DT 13-OCT-1998 (first entry)
DE HPIV-3 FRH1 cp45 vaccine L protein.
KW L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KW single stranded RNA virus; Mononegavirales.
OS Human parainfluenza virus.
PN WO9813501-A2.
PD 02-APR-1998.
PF 19-SEP-1997; U16718.
PR 27-SEP-1996; US-026823.
PA (AMCY ) AMERICAN CYANAMID CO.
PT (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
DR WPI: 98-230710/20.
N-PSDB: V18274\..
DR N-PSDB: V18273\..
PT Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT single stranded RNA virus of order Mononegavirales - having
PT attenuating mutation in 3' genomic promoter region and RNA
PT polymerase gene, useful as vaccine to immunise against such virus
PS Disclosure: Page 265-273; 426pp; English.
This sequence represents the human parainfluenza virus (HPIV-3) type 3
vaccine FRH1 cp45 L protein. This sequence is used in a method which
CC involves the isolation of recombinantly-generated, attenuated,
CC non-segmented, negative-sense, single stranded RNA virus of the order
CC Mononegavirales which have at least 1 attenuating mutation in the 3'
CC genomic promoter region and at least 1 attenuating mutation in the RNA
CC polymerase gene. This RNA virus can be used as a vaccine to immunise an
CC individual against such a virus.
SQ Sequence 2233 AA;

Query Match          56.3%; Score 54; DB 33; Length 2233;
Best Local Similarity 33.3%; Pred. No. 4.45e+01;
Matches      5; Conservative       7; Mismatches    3; Indels     0; Gaps     0;

Db      973 ldrsvlyrimnqepg 987
      | : : : : | : ||| :
Oy      1 LNSKIAFRIVSOEPA 15

RESULT   13
ID W48713 standard; Protein; 2233 AA.
AC W48713;
DT 13-OCT-1998 (first entry)
DE HPIV-3 Vero cp45 vaccine L protein.
KW L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KW single stranded RNA virus; Mononegavirales.
OS Human parainfluenza virus.
PN WO9813501-A2.
PD 02-APR-1998.
PF 19-SEP-1997; U16718.
PR 27-SEP-1996; US-026823.
PA (AMCY ) AMERICAN CYANAMID CO.
PT (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
DR WPI: 98-230710/20.
N-PSDB: V18274.
DR Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT single stranded RNA virus of order Mononegavirales - having
```

```

PT attenuating mutation in 3' genomic promoter region and RNA
PT polymerase gene, useful as vaccine to immunise against such virus
PS Disclosure, Page 283-291, 426pp, English.
CC This sequence represents the Human parainfluenza virus (HPIV-3) type 3
CC vaccine Vero cp45 L protein. This sequence is used in a method which
CC involves the isolation of recombinantly-generated, attenuated,
CC non-segmented, negative-sense, single stranded RNA virus of the order
CC Mononegavirales which have at least 1 attenuating mutation in the 3'
CC genomic promoter region and at least 1 attenuating mutation in the RNA
CC polymerase gene. This RNA virus can be used as a vaccine to immunise an
CC individual against such a virus.
SQ Sequence 2233 AA:

Dd 973 ldrsvlyymqpepg 987
Qy 1 LNSKIAKIVSQEPA 15

RESULT 14
ID W82841 standard; Protein; 456 AA.
AC W82841;
DE 03-FEB-1999 (first entry)
DE Human cerebral protein-1.
DE Human; cerebral protein-1; HUCEP-1; ischaemic brain disease;
DE neuron function activating activity; nerve denatured disease;
KW Parkinson's disease; Alzheimer's disease.
OS Homo sapiens.
PN J10257891-A.
PD 29-SEP-1998.
PE 19-MAR-1997; 065716.
PR 19-MAR-1997; JP-065716.
PA (TAIS ) TAISHO PHARM CO LTD.
DR WPI: 98-575902/49.
DR N-PSDB: V64182.
PT HUCEP-1 protein having neuron function activating activity - useful
PT for treating ischaemic brain diseases and nerve denaturation
PT conditions such as Parkinson's and Alzheimer's diseases
PS Claim 1; Fig 4; 17pp; Japanese.
CC The present sequence represents human cerebral protein-1 (HUCEP-1).
CC HUCEP-1 has neuron function activating activity. HUCEP-1 is useful
CC for treating ischaemic brain diseases and nerve denaturation
CC conditions such as Parkinson's and Alzheimer's diseases.
SQ Sequence 456 AA;

Query Match 55.2%; Score 53; DB 37; Length 456;
Best Local Similarity 54.5%; Pred. No. 5.74e+01;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0.

Dd 408 tklgfkkyvskd 418
Qy 3 SKIAKIVSQE 13

RESULT 15
ID R40843 standard; Protein; 572 AA.
AC R40843;
DE 24-FEB-1994 (first entry)
DE Bilirubin oxidase.
DE Bilirubin; oxidase; analytical; BO; expression vector; PCR;
KW polymerase chain reaction.
OS Myrothecium verrucaria.
FH Key Location/Qualifiers
FT peptide 1..38
FT /label= sig_peptide
FT 39..534
FT /label= mat_protein

J05199882-A.
PD 10-AUG-1993.
PD 24-JAN-1992; 034126.

```




Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Merch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Jun 11 17:25:31 1999; MasPar time 4.28 Seconds
Molecular output not generated.

Title: >US-08-991-628-3
Description: (1-15) from US08991628.pep
Perfect score: 96
Sequence: 1 LNSKRAFRIVSQEPA 15

Scoring table: PAM 150
Gap 15

Searched: 122810 seqs, 40068593 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r60
1: p1r1 2: p1r2 3: p1r3 4: p1r4

Statistics: Mean 25.594; Variance 33.108; scale 0.773

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	96	100.0	999	1 IJHUG3	desmoglein 3 precursor	2.05e-09
2	87	90.6	1043	1 IJHUG1	desmoglein 1 precursor	3.51e-07
3	87	90.6	1049	1 IJHUG1	desmoglein 1 precursor	3.51e-07
4	77	80.2	1117	2 S38673	desmoglein 2 - human	8.47e-05
5	61	63.5	1421	2 T02501	hypothetical protein	2.82e-01
6	60	62.5	884	1 IJMSCB	E-cadherin precursor	4.53e-01
7	60	62.5	884	1 IJMSCB	E-cadherin precursor	4.53e-01
8	60	62.5	906	1 IJHUCN	cadherin 2 precursor	4.53e-01
9	59	61.5	385	2 S49752	homeotic protein YOX1	7.23e-01
10	58	60.4	138	2 S74520	hypothetical protein	1.15e+00
11	57	59.4	1264	2 S64146	probable membrane pro	1.81e+00
12	57	59.4	5147	1 IJFFRM	cadherin-related tumo	4.46e+00
13	55	57.3	192	2 E70126	transactin elongatio	6.93e+00
14	54	56.3	877	1 IJBOCN	N-cadherin precursor	6.93e+00
15	54	56.3	882	1 IJHUCB	cadherin 1 precursor	6.93e+00
16	54	56.3	905	2 S43064	cadherin - African cl	6.93e+00
17	54	56.3	906	2 IJMSCN	N-cadherin precursor	6.93e+00
18	54	56.3	1146	2 S64402	probable transcriptio	6.93e+00
19	54	55.2	2233	1 ZLNZP3	genome polyprotein -	6.93e+00
20	53	55.2	887	1 IJHUCB	E-cadherin precursor	1.07e+01
21	53	55.2	2048	1 ZLNZSE	genome polyprotein -	1.07e+01
22	53	55.2	2228	1 ZLNZSE	genome polyprotein -	1.07e+01
23	52	54.2	485	2 A71266	probable transcriptio	1.65e+01

24	52	54.2	732	1 IJHUCB	B-cadherin precursor	1.65e+01
25	52	54.2	821	2 B34488	calpain (EC 3.4.22.17	1.65e+01
26	51	53.1	241	2 S63634	ribosomal protein S3	2.51e+01
27	51	53.1	293	2 F49846	spaz3 protein - Shige	2.51e+01
28	51	53.1	293	2 E42284	spaz3 protein - Shige	2.51e+01
29	51	53.1	644	2 G64938	hypothetical protein	2.51e+01
30	51	53.1	742	2 A49341	isocitrate dehydrogen	2.51e+01
31	51	53.1	813	3 T02672	hypothetical protein	2.51e+01
32	50	52.1	154	2 B70351	ribosomal-protein-ala	3.81e+01
33	50	52.1	194	2 I51310	beta 8 integrin - chi	3.81e+01
34	50	52.1	217	2 S25314	aspartic proteinase 1	3.81e+01
35	50	52.1	261	2 H69779	antibiotic resistance	3.81e+01
36	50	52.1	276	2 JC5285	carboxyl reductase (N	3.81e+01
37	50	52.1	277	2 JC5284	carboxyl reductase (N	3.81e+01
38	50	52.1	461	2 A71662	sodium/pantothenate s	3.81e+01
39	50	52.1	506	2 G69546	hypothetical protein	3.81e+01
40	50	52.1	578	2 T02292	hypothetical protein	3.81e+01
41	50	52.1	2410	1 JQ1948	genome polyprotein 1	3.81e+01
42	50	52.1	2412	1 JQ1537	genome polyprotein 1	3.81e+01
43	50	52.1	308	2 E71006	probable GMP synthase	5.75e+01
44	49	51.0	342	1 CFPM	plastoquinol-1-plasto	5.75e+01
45	49	51.0	1715	2 JE0128	Bombyx mori receptor	5.75e+01

ALIGNMENTS

RESULT 1
ENTRY IJHUG3 #type complete
TITLE desmoglein 3 precursor - human
ALTERNATE_NAMES pemphigus vulgaris antigen
ORGANISM Homo sapiens #common_name man
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 18-Sep-1998

ACCESSIONS
REFERENCE A41088
#authors Amagai, M.; Klaus-Kovtun, V.; Stanley, J.R.
#journal Cell (1991) 67:869-877
#title Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a disease of cell adhesion.

#cross-references NID:92069753
#accession A41088
##molecule_type mRNA
##residues 1-999 #label AMA
##cross-references GB:M76482; NID:9190751; PID:9190752

GENETICS

#gene GDB:DSG3
##cross-references GDB:134030; OMIM:169615

CLASSIFICATION #map_position 18q12.1-18q12.2
KEYWORDS #superfamily cadherin; cadherin repeat homology calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein

FEATURE
1-23 #domain signal sequence #status predicted #label SIG
24-49 #domain propeptide #status predicted #label PRO
50-999 #product desmoglein homolog #status predicted #label MAR

50-615 #domain extracellular #status predicted #label EXT
52-157 #domain cadherin repeat homology #label CR1
160-267 #domain cadherin repeat homology #label CR2
270-493 #domain cadherin repeat homology #label CR3
390-495 #domain cadherin repeat homology #label CR5
416-598 #domain cadherin repeat homology #label CR5
616-639 #domain transmembrane #status predicted #label TM
640-999 #domain intracellular #status predicted #label INT
910-938 #domain desmoglein repeat #label DGL
937-966 #domain desmoglein repeat #label DGL
110,180,545 #domain desmoglein repeat #label DGL
binding_site carboxylate (Asn) (covalent) #status predicted

SUMMARY #length 999 #molecular-weight 107502 #checksum 8311
Query Match 100.0%; Score 96; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 2.05e-09;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 190 LNSKIAFKIVSOEPA 204
 1 LNSKIAFKIVSOEPA 15
 RESULT 2
 ENTRY 1BBOGI #type complete
 TITLE desmoglein 1 precursor - bovine
 ORGANISM BDM
 #formal_name BDM
 #sequence_revision 30-Jun-1993 #text_change 05-Sep-1997
 ACCESSIONS S14603; A38872; A37785; S38721; A48173; S24412
 REFERENCE S14603; A38872; A37785; S38721; A48173; S24412
 #authors Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
 #description Submitted to the EMBL Data Library, March 1991
 Complete sequence of the desmoglein precursor and evidence for the existence of different desmoglein genes expressed in cell type-specific patterns.
 #accession S14603
 #molecule_type mRNA
 #residues 1-1043 #label KOC
 REFERENCE #cross-references EMBL:X58466; NID:9306; PID:9307
 A38872
 #authors Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.
 #journal Eur. J. Cell Biol. (1991) 55:200-208
 #title Complete amino acid sequence of the epidermal desmoglein precursor polypeptide and identification of a second type of desmoglein gene.
 #cross-references M01D:92037656
 #accession A38872
 #molecule_type mRNA
 #residues 1-87:968-1043 #label KO2
 REFERENCE #cross-references GB:S64268; GB:S64270
 A37785
 #authors Goodwin, L.; Hill, J.E.; Raynor, K.; Raszi, L.; Manabe, M.; Cowin, P.
 #journal Biochem. Biophys. Res. Commun. (1990) 173:1224-1230
 #title Desmoglein shows extensive homology to the cadherin family of cell adhesion molecules.
 #cross-references M01D:91097553
 #accession A37785
 #molecule_type mRNA
 #residues 44-123, 'V', 125-493 #label GOO
 REFERENCE #cross-references GB:M58165; NID:9162966; PID:9552318
 S38721
 #authors Zimbelmann, R.
 #submission Submitted to the EMBL Data Library, February 1991
 #accession S38721
 #molecule_type mRNA
 #residues 44-1043 #label ZIM
 REFERENCE #cross-references EMBL:X57784; NID:9436061; PID:9436062
 A48173
 #authors Koch, P.J.; Walsh, M.J.; Schmelz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
 #journal Eur. J. Cell Biol. (1990) 53:1-12
 #title Identification of desmoglein, a constitutive desmosomal glycoprotein, as a member of the cadherin family of cell adhesion molecules.
 #cross-references M01D:91168965
 #accession A48173
 #molecule_type mRNA
 #residues 44-1001, 'AOPPSAT' #label KO3
 REFERENCE #cross-references GB:X57784
 #note This sequence has been revised in references A38872 and S38721
 GENETICS
 #gene DSG1
 #classification #superfamily cadherin; cadherin repeat homology
 #keywords calcium binding; cell adhesion; duplication; glycoprotein; Transmembrane protein

FEATURE
 1-23
 24-49
 50-1043
 50-548
 52-157
 160-269
 272-385
 392-491
 549-574
 575-1043
 846-875
 876-905
 906-933
 934-962
 963-1012
 110
 180,496
 SUMMARY #length 1043 #molecular_weight 112242 #checksum 6897
 Query Match 90.6%; Score 87; DB 1; Length 1043;
 Best Local Similarity 80.0%; Pred. No. 3.51e-07;
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 DB 190 LNSKIAFKIVSOEPA 204
 1 LNSKIAFKIVSOEPA 15
 RESULT 3
 ENTRY 1BBOGI #type complete
 TITLE desmoglein 1 precursor - human
 ORGANISM HOMO
 #formal_name Homo sapiens #common_name man
 #sequence_revision 30-Jun-1993 #text_change 26-Feb-1998
 ACCESSIONS S16906; A39706; A61254; A61279; S16158
 REFERENCE S16906; A39706; A61254; A61279; S16158
 #authors Buxton, R.S.
 #submission Submitted to the EMBL Data Library, November 1990
 #accession S16906
 #molecule_type mRNA
 #residues 1-1049 #label BUX
 REFERENCE #cross-references EMBL:X56654; NID:930505; PID:930506
 A39706
 #authors Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Ataliotis, P.; Poynter, D.; Arnemann, J.; Rutman, A.J.; Pilsley, S.C.; Watt, F.M.; Rees, D.A.; Buxton, R.S.; Magee, A.I.
 #journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:4796-4800
 #title Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions, is related to the cadherin family of cell adhesion molecules.
 #cross-references M01D:91271279
 #accession A39706
 #molecule_type mRNA
 #residues 24-1049 #label WHE
 REFERENCE #cross-references GB:X56654
 A61254
 #authors Nilles, L.A.; Parry, D.A.D.; Powers, E.E.; Angst, B.D.; Wagner, R.M.; Green, K.J.
 #journal J. Cell Sci. (1991) 99:809-821
 #title Structural analysis and expression of human desmoglein: a cadherin-like component of the desmosome.
 #accession A61254
 #molecule_type mRNA
 #residues 26-1049 #label NIL
 REFERENCE A61279
 #authors Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arnemann, J.; Rees, D.A.; King, I.A.; Magee, A.I.
 #journal Biochem. Soc. Trans. (1991) 19:1060-1064
 #title Desmosomal glycoproteins I, II and III: novel members of the

```
cadherin superfamily.
#cross-references MUID:92175187
#accession A61279
#status not compared with conceptual translation
#molecule-type mRNA
#residues 1-55 #label WH3

GENETICS
#gene GDB:DSG1
#cross-references GDB:126563; OMTM:125670
#map_position 18q12.1-18q12.2
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;
transmembrane protein

FEATURE
1-23
24-49 #domain signal sequence #status predicted #label SIG\
50-1049 #domain propeptide #status predicted #label PRO\
50-548 #product desmoglein #status predicted #label MAR\
52-157 #domain extracellular #status predicted #label EXT\
160-269 #domain cadherin repeat homology #label CR1\
272-385 #domain cadherin repeat homology #label CR2\
392-493 #domain cadherin repeat homology #label CR3\
509-550 #domain cadherin repeat homology #label CR4\
549-569 #region serine/threonine-rich\
572-1049 #domain transmembrane #status predicted #label TM\
840-869 #domain intracellular #status predicted #label INT\
870-899 #domain desmoglein repeat #label D61\
900-927 #domain desmoglein repeat #label D62\
928-956 #domain desmoglein repeat #label D63\
969-1019 #region glycine/serine-rich\
110,180 #binding-site carbohydrate (Asn) (covalent) #status
predicted

SUMMARY #length 1049 #molecular-weight 113715 #checksum 4482

Query Match 90.6%; Score 87; DB 1; Length 1049;
Best Local Similarity 80.0%; Pred. No. 3,51e-07;
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 190 LNSKIAFKIRPEPS 204
QY 1 LNSKIAFKIVSOEPA 15

RESULT 4
ENTRY S38673 #type complete
TITLE desmoglein 2 - human
ALTERNATE_NAMES desmoglein HDGC
ORGANISM #formal_name Homo sapiens #common_name man
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
08-Sep-1997
ACCESSIONS S38673; B38872
REFERENCE S38673
#authors Zimbelmann, R.
#submission submitted to the EMBL Data Library, September 1993
#accession S38673
#status preliminary
#molecule-type mRNA
#residues 1-1117 #label ZIM
#cross-references EMBL:Z26317; NID:g416177; PID:g416178
A38872
#authors Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.;
Frankel, W.W.
#journal Eur. J. Cell Biol. (1991) 55:200-208
#title Complete amino acid sequence of the epidermal desmoglein
precursor polypeptide and identification of a second type
of desmoglein gene.
#cross-references MUID:92037656
#accession B38872
#molecule-type mRNA
#residues 777-1117 #label KOC
#cross-references GB:S64273

GENETICS
#gene GDB:DSG2

cadherin superfamily.
#cross-references GDB:128808; OMTM:125671
#map_position 18q12.1-18q12.2
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;
membrane protein

FEATURE
51-158
161-271 #domain cadherin repeat homology #label CR1\
#domain cadherin repeat homology #label CR2\
#length 1117 #molecular-weight 122384 #checksum 7660

SUMMARY
Query Match 80.2%; Score 77; DB 2; Length 1117;
Best Local Similarity 73.3%; Pred. No. 8,47e-05;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 191 LNSKISYRIVSLEPA 205
QY 1 LNSKIAFKIVSOEPA 15

RESULT 5
ENTRY T02501 #type complete
TITLE hypothetical protein T19C21.7 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change
05-Mar-1999
ACCESSIONS T02501
REFERENCE Z14676
#authors Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.;
Brandon, R.C.; Sykes, S.M.; Kail, S.; Mason, T.M.;
Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter,
J.C.
#submission submitted to the EMBL Data Library, August 1998
#description Arabidopsis thaliana chromosome II BAC T19C21 genomic
sequence.
#accession T02501
#status preliminary; translated from GB/EMBL/DBJ
#molecule-type DNA
#residues 1-1421 #label R07
#cross-references EMBL:AC004683; NID:g3395421; PID:g3395428

GENETICS
#map_position 2
#introns 52/2; 107/1; 148/2; 191/3; 212/2; 1317/3; 1346/3
#note T19C21.7
SUMMARY #length 1421 #molecular-weight 154325 #checksum 9888

Query Match 63.5%; Score 61; DB 2; Length 1421;
Best Local Similarity 40.0%; Pred. No. 2,82e-01;
Matches 6; Conservative 8; Mismatches 1; Indels 0; Gaps 0;

Db 262 VDSKISYEITONG 276
QY 1 LNSKIAFKIVSOEPA 15

RESULT 6
ENTRY IIMSCE #type complete
TITLE E-cadherin precursor, epithelial - mouse
ALTERNATE_NAMES uvomorulin
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
13-Mar-1997
ACCESSIONS S04528; S03160; I49565; S48735
REFERENCE S04528
#authors Nagafuchi, A.; Shirayoshi, Y.; Okazaki, K.; Yasuda, K.;
Takahashi, M.
#journal Nature (1987) 329:341-343
#title Transformation of cell adhesion properties by exogenously
introduced E-cadherin cDNA.
#cross-references MUID:87315445
#accession S04528
#molecule-type mRNA
#residues 1-412, 'V', 414-884 #label NAG
```

```

#cross-references EMBL:X06115
REFERENCE
#authors Ringwald, M.; Schuh, R.; Vestweber, D.; Eistetter, H.;
#journal Lottspeich, F.; Engel, J.; Doeltz, R.; Jaehning, F.; Epplen,
#title J. Mayer, S.; Mueller, C.; Kemler, R.
#title EMO J. (1987) 6:3647-3653
The structure of cell adhesion molecule uvomorulin. Insights
into the molecular mechanism of Ca(2+)-dependent cell
adhesion.
#cross-references MWID:8811553
#accession S03160
#molecule-type mRNA
#residues 157-884 #label RIN
#cross-references EMBL:X06339
#note part of this sequence, including the amino end of the
mature protein, was confirmed by protein sequencing
REFERENCE
#authors Behrens, J.; Loewrick, O.; Klein-Hitpass, L.; Birchmeier, W.
#journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:11495-11499
#title The E-cadherin promoter: Functional analysis of a G-C-rich
region and an epithelial cell-specific palindromic
regulatory element.
#cross-references MWID:92107977
#accession I49565
#status preliminary; translated from GB/EMBL/DBJ
#molecule-type DNA
#residues 1-15 #label RES
#cross-references GB:M61449; NID:g192325; PID:g192326
REFERENCE
#authors Tong, K.I.; Yau, P.; Overduin, M.; Bagby, S.; Porumb, T.;
#journal Takeichi, M.; Ikura, M.
#title FEBS Lett. (1994) 352:318-322
Purification and spectroscopic characterization of a
recombinant amino-terminal polypeptide fragment of mouse
epithelial cadherin.
#cross-references MWID:95010732
#accession S48735
#status preliminary
#molecule-type protein
#residues 156-300 #label TON
COMMENT Cadherins mediate calcium-dependent intercellular adhesion, and are
thought to be involved in the sorting of different cell types
during morphogenesis.
GENETICS
#gene E-cadherin
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;
transmembrane protein
FEATURE
1-27 #domain signal sequence #status predicted #label SIG\
28-156 #domain propeptide #status predicted #label PRO\
157-884 #product E-cadherin, epithelial #status experimental
#label MAR\
157-699 #domain extracellular #status predicted #label EXT\
159-264 #domain cadherin repeat homology #label CR1\
234-239 #region cadherin binding #status predicted\
267-377 #domain cadherin repeat homology #label CR2\
380-488 #domain cadherin repeat homology #label CR3\
489-597 #domain cadherin repeat homology #label CR4\
598-702 #domain cadherin repeat homology #label CR5\
702-733 #domain transmembrane #status predicted #label TM\
734-884 #domain intracellular #status predicted #label INT\
842-855 #region serine-rich\
560,639 #binding_site carboxylate (Asn) (covalent) #status
predicted
SUMMARY #length 884 #molecular-weight 98255 #checksum 2268
Query Match 62.5%; Score 60; DB 1; Length 884;
Best Local Similarity 61.5%; Pred. No. 4.53e-01;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
DB 299 NAIATVVSQDP 311
I: ||: ||||: |

```

```

QY 2 NSKIAFKIVSQEP 14
RESULT 7
ENTRY S34438 #type complete
TITLE uvomorulin - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 22-Nov-1993 #sequence_revision 03-Nov-1995 #text_change
23-May-1997
ACCESSIONS S34438
REFERENCE S34438
#authors Ringwald, M.; Baribault, H.; Schmidt, C.; Kemler, R.
#journal Nucleic Acids Res. (1991) 19:6533-6539
#title The structure of the gene coding for the mouse cell adhesion
molecule uvomorulin.
#cross-references MWID:92093614
#accession S34438
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule-type DNA
#residues 1-884 #label RIN
#cross-references EMBL:X06975
#note the nucleotide sequence was submitted to the EMBL Data
Library, August 1992
GENETICS
#introns 16/3; 57/1; 131/3; 179/3; 231/3; 280/1; 338/3; 381/3; 442/3;
524/2; 573/1; 648/1; 724/1; 767/3; 815/3
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
FEATURE
SUMMARY #domain cadherin repeat homology #label CR3
#length 884 #molecular-weight 98283 #checksum 2125
Query Match 62.5%; Score 60; DB 2; Length 884;
Best Local Similarity 61.5%; Pred. No. 4.53e-01;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
DB 299 NAIATVVSQDP 311
I: ||: ||||: |
QY 2 NSKIAFKIVSQEP 14
RESULT 8
ENTRY IOHUCN #type complete
TITLE cadherin 2 precursor - human
ALTERNATE_NAMES N-cadherin; neuronal cadherin
ORGANISM #formal_name Homo sapiens #common_name man
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
26-Feb-1998
ACCESSIONS A38870; S11487; J00751; S13799
REFERENCE A38870
#authors Reid, R.A.
#submission submitted to the EMBL Data Library, November 1990
#accession A38870
#molecule-type mRNA
#residues 1-906 #label REI
#cross-references EMBL:X54315; NID:g34998; PID:g34999
REFERENCE S11487
#authors Reid, R.A.; Hemperly, J.J.
#journal Nucleic Acids Res. (1990) 18:5896
#title Human N-cadherin: nucleotide and deduced amino acid sequence.
#cross-references MWID:91016946
#accession S11487
#molecule-type mRNA
#residues 1-340, 'N', 342-698, 'R', 700-704, 'F', 706-906 #label RE2
#cross-references EMBL:X54315
#note this sequence has been revised in reference A38870
REFERENCE J00751
#authors Walsh, F.S.; Barton, C.H.; Putt, W.; Moore, S.E.; Kelsell,
#journal D.; Spurr, N.; Goodfellow, P.N.
#title J. Neurochem. (1990) 55:805-812
N-cadherin gene maps to human chromosome 18 and is not linked
to the E-cadherin gene.
#cross-references MWID:90347462

```

```

#accession J00751
#molecule_type mRNA
#residues 160-194, 'IR', 197-211, 'L', 213-227, 'Q', 229, 'N', 231-235,
#237-248, 'T', 250-356, 'N', 358-530, 'KYL', 533-906
#label MAL
##cross-references GB:M34064
COMMENT Cadherins mediate calcium-dependent intercellular adhesion and are
thought to be involved in the sorting of different cell types
during morphogenesis.

GENETICS
#gene GDB:CDH2; NCAD
##cross-references GDB:128185; OMIM:114020
#map position 18q12.1-18q12.1
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;
transmembrane protein

FEATURE
1-27 #domain signal sequence #status predicted #label SIG
28-159 #domain propeptide #status predicted #label PRO
160-906 #product N-cadherin #status predicted #label MAT
160-714 #domain extracellular #status predicted #label EXT
162-267 #domain cadherin repeat homology #label CRI
237-242 #region cadherin binding #status predicted
270-382 #domain cadherin repeat homology #label CR2
385-497 #domain cadherin repeat homology #label CR3
500-605 #domain cadherin repeat homology #label CR4
606-712 #domain cadherin repeat homology #label CR5
715-746 #domain transmembrane #status predicted #label TMN
747-906 #domain intracellular #status predicted #label INT
865-878 #region serine-rich
150, 273, 325, 402,
572, 622, 651, 692,
#binding_site carbohydrate (asn) (covalent) #status
predicted
SUMMARY #length 906 #molecular-weight 99864 #checksum 9191

Query Match 62.5%; Score 60; DB 1; Length 906;
Best Local Similarity 46.7%; Pred. NO. 4,53e-01;
Matches 7; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Db 300 LNCMARRIVSQAPS 314
1 LNSKIARIVSQEPA 15

RESULT 9
ENTRY S49752 #type complete
TITLE homeotic protein YOX1 - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES #formal_name Saccharomyces cerevisiae
ORGANISM 13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change
DATE 06-Feb-1998
ACCESSIONS S49752; S33388
REFERENCE S49741
#authors Badcock, K.; Churcher, C.
#submission Submitted to the EMBL Data Library, November 1994
#accession S49752
#molecule_type DNA
#residues 1-385 #label BAD
##cross-references EMBL:246659; NID:g575680; PID:g575692; MIPS:YML027*
REFERENCE S33388
#authors Kautmann, E.
#journal Chromosoma (1993) 102:174-179
#title In vitro binding to the leucine tRNA gene identifies a novel
yeast homeobox gene.
#cross-references MIM:93209080
#accession S33388
#molecule_type DNA
#residues 1-290, 'OGIIP', #label KAU
##cross-references EMBL:X62392; NID:g5501; PID:g5502
GENETICS SGD:YOX1
#gene #cross-references SGD:S0004489; MIPS:YML027*
#map_position 13L

```

```

CLASSIFICATION #superfamily unassigned homeobox proteins; homeobox homology
KEYWORDS DNA binding; homeobox; nucleus; transcription regulation
FEATURE 177-233 #domain homeobox homology #label HOX
SUMMARY #length 385 #molecular-weight 42739 #checksum 6962

Query Match 61.5%; Score 59; DB 2; Length 385;
Best Local Similarity 63.6%; Pred. NO. 7.23e-01;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 364 KGFRTVDOCP 374
4 KIAFRIVSQEP 14

RESULT 10
ENTRY S74520 #type complete
TITLE hypothetical protein slr0725 - Synechocystis sp. (strain PCC
6803)
ORGANISM #formal_name Synechocystis sp.
DATE #variety
25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
ACCESSIONS S74520
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
S.; Takeuchi, C.; Mada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular
cyanobacterium Synechocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references MIM:97061201
#accession S74520
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-138 #label KAN
##cross-references EMBL:D90899; GB:AB001339; NID:g1651650; PID:d1017405;
PID:g1651744
#note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996

GENETICS
#start_codon GTG
CLASSIFICATION #superfamily hypothetical protein slr0725
SUMMARY #length 138 #molecular-weight 16580 #checksum 4882

Query Match 60.4%; Score 58; DB 2; Length 138;
Best Local Similarity 38.5%; Pred. NO. 1.15e+00;
Matches 5; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db 28 AKLAFVTPDPT 40
3 KIAFRIVSQEPA 15

RESULT 11
ENTRY S64146 #type complete
TITLE probable membrane protein YGL133w - yeast (Saccharomyces
cerevisiae)
ALTERNATE_NAMES #formal_name Saccharomyces cerevisiae
ORGANISM 17-May-1998 #sequence_revision 17-May-1998 #text_change
DATE 17-Apr-1998
ACCESSIONS S64146; S71739
REFERENCE S64144
#authors Escrichano, V.; Erraso, P.; Portillo, F.; Mazon, M.J.
#submission Submitted to the Protein Sequence Database, May 1996
#accession S64146
#molecule_type DNA

```

```

REFERENCE
#authors Escibano, V.; Erraso, P.; Portillo, F.; Mazon, M.J.
#journal Yeast (1996) 12:887-892
#title Sequence analysis of a 14.6 kb DNA fragment of Saccharomyces
cerevisiae chromosome VII reveals SEC27, SSMb, a putative
S-adenosylmethionine-dependent enzyme and six new open
reading frames.
#cross-references MIM:26437978
#accession S71739
#status nucleic acid sequence not shown; translation not shown
#molecule-type DNA
#residues 1-1264 #label ESM
#cross-references EMBL:X92670
#note the nucleotide sequence was submitted to the EMBL Data
Library, October 1995

GENETICS
#map-position 7L
#feature Transmembrane protein
FEATURE
#feature 549-565
#length 1264 #molecular-weight 145642 #checksum 9929
SUMMARY
Query Match 59.4%; Score 57; DB 2; Length 1264;
Best Local Similarity 42.9%; Pred. No. 1.81e+00;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 NSKIAFRIVSQEPA 15
SDKVEIKIVHDDPA 537
: 1: 111 111
: 2: 111 111

RESULT 12
ENTRY IUFFTM #type complete
#feature cadherin-related tumor suppressor precursor - fruit fly
#formal-name Drosophila melanogaster
ORGANISM #formal-name Drosophila melanogaster
DATE 30-Sep-1993 #sequence-revision 30-Sep-1993 #text-change
16-Feb-1997
ACCESSIONS A41087; B41087
REFERENCE A41087
#authors Mahoney, P.A.; Weber, U.; Onofrechuk, P.; Biesmann, H.;
Bryant, P.J.; Goodman, C.S.
#journal Cell (1991) 67:853-868
#title The fat tumor suppressor gene in Drosophila encodes a novel
member of the cadherin gene superfamily.
#cross-references MIM:92069752
#accession A41087
#molecule-type mRNA
#residues 143-485:1279-5147 #label MAH
#cross-references GB:M80537
#accession B41087
#molecule-type DNA
#residues 1-142:487-1278 #label MA2
#cross-references GB:M80537
#note 1229-Gly and 1233-Ser were also found

GENETICS
#gene fat
#cross-references FlyBase:FBgn0001075
CLASSIFICATION #superfamily cadherin-related tumor suppressor; cadherin
repeats; EGF homology
KEYWORDS calcium binding; cell adhesion; duplication; transmembrane
protein
FEATURE
#feature 1-35
#domain signal sequence #status predicted #label SIG\
36-5147 #product cadherin-related tumor suppressor #status
predicted #label MARY\
51-156 #domain extracellular #status predicted #label CR1\
159-270 #domain cadherin repeat homology #label CR2\
#domain cadherin repeat homology #label CR2\

```

```

271-382 #domain cadherin repeat homology #label CR3\
390-494 #domain cadherin repeat homology #label CR4\
497-599 #domain cadherin repeat homology #label CR5\
602-708 #domain cadherin repeat homology #label CR6\
718-822 #domain cadherin repeat homology #label CR7\
831-942 #domain cadherin repeat homology #label CR8\
948-1049 #domain cadherin repeat homology #label CR9\
1052-1153 #domain cadherin repeat homology #label CR10\
1156-1278 #domain cadherin repeat homology #label CR11\
1281-1384 #domain cadherin repeat homology #label CR12\
1387-1489 #domain cadherin repeat homology #label CR13\
1492-1601 #domain cadherin repeat homology #label CR14\
1607-1713 #domain cadherin repeat homology #label CR15\
1717-1823 #domain cadherin repeat homology #label CR16\
1826-1922 #domain cadherin repeat homology #label CR17\
1925-2027 #domain cadherin repeat homology #label CR18\
2028-2167 #domain cadherin repeat homology #label CR19\
2169-2278 #domain cadherin repeat homology #label CR20\
2281-2384 #domain cadherin repeat homology #label CR21\
2387-2491 #domain cadherin repeat homology #label CR22\
2494-2596 #domain cadherin repeat homology #label CR23\
2599-2703 #domain cadherin repeat homology #label CR24\
2707-2810 #domain cadherin repeat homology #label CR25\
2813-2913 #domain cadherin repeat homology #label CR26\
2915-3014 #domain cadherin repeat homology #label CR27\
3014-3124 #domain cadherin repeat homology #label CR28\
3127-3229 #domain cadherin repeat homology #label CR29\
3232-3334 #domain cadherin repeat homology #label CR30\
3337-3439 #domain cadherin repeat homology #label CR31\
3442-3545 #domain cadherin repeat homology #label CR32\
3548-3651 #domain cadherin repeat homology #label CR33\
3654-3756 #domain cadherin repeat homology #label CR34\
3954-4010 #domain cadherin repeat homology #label CR35\
4017-4048 #domain EGF homology #label EG1\
4056-4089 #domain EGF homology #label EG2\
4096-4127 #domain EGF homology #label EG3\
4584-4609 #domain EGF homology #label EG4\
4610-5147 #domain intracellular #status predicted #label INT\
#domain intracellular #status predicted #label INT\
#length 5147 #molecular-weight 564895 #checksum 6994

SUMMARY
Query Match 59.4%; Score 57; DB 1; Length 5147;
Best Local Similarity 50.0%; Pred. No. 1.81e+00;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 1520 LNKVSYAISKQEP 1533
QY 1 LNSKIAFRIVSQEP 14
||:|::| |||
||:|::| |||

RESULT 13
ENTRY F70126 #type complete
#feature translocation elongation factor P (efp) homolog - Lyme disease
spirochete
#formal-name Borrelia burgdorferi #common-name Lyme disease
spirochete
ORGANISM #formal-name Borrelia burgdorferi
DATE 13-Sep-1998 #sequence-revision 13-Feb-1998 #text-change
13-Sep-1998
ACCESSIONS F70126
REFERENCE A70100
#authors Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.;
Clayton, R.; Lathigra, R.; White, O.; Dougherty, B.; Tomb,
Dodson, R.; Hickley, E.K.; Gwinn, M.; Dougherty, B.; Tomb,
J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.;
Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
Vogt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman,
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst,
K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
#journal Nature (1997) 390:580-586
#title Genomic sequence of a Lyme disease spirochete, Borrelia
burgdorferi.
#cross-references MIM:98065943
#accession F70126

```

```
##status preliminary; nucleic acid sequence not shown;
##molecule_type DNA translation not shown.
##residues 1-192 ##label KLE
##cross-references GB:AE001132; GB:AE000783; NID:g2688107; PID:g2688116;
TIGR:BB0214
##experimental source strain B31
##restriction superfamily translation elongation factor EF-P
SUMMARY #length 192 #molecular-weight 21413 #checksum 4881

Query Match 57.3%; Score 55; DB 2; Length 192;
Best Local Similarity 53.3%; Pred. No. 4.46e+00;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Lb 130 LAPKAFVEVEEA 144
Oy 1 LNSKIAFKIVSOEPA 15

RESULT 14
ENTRY IUBCN #type fragment
TITLE N-cadherin precursor - bovine (fragment)
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
05-Sep-1997
ACCESSIONS S11693
REFERENCE S11693
#authors Liaw, C.W.; Cannon, C.; Power, M.D.; Kiboneka, P.K.; Rubin, L.L.
#journal EXMO J. (1990) 9:2701-2708
#title Identification and cloning of two species of cadherins in bovine endothelial cells.
#cross-references M01D:90360979
#accession S11693
##molecule_type mRNA
##residues 1-877 ##label LIA
##cross-references EMBL:X53615; NID:g164; PID:g664894
COMMENT ##cross-references EMBL:X53615; NID:g164; PID:g664894
Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in the sorting of different cell types during morphogenesis.
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein
FEATURE
1-130 #domain propeptide (fragment) #status predicted #label PRO\
131-877 #product N-cadherin #status predicted #label MAN\
131-685 #domain extracellular #status predicted #label EXT\
133-238 #domain cadherin repeat homology #label CR1\
208-213 #region cadherin binding #status predicted\
241-353 #domain cadherin repeat homology #label CR2\
356-468 #domain cadherin repeat homology #label CR3\
471-576 #domain cadherin repeat homology #label CR4\
577-685 #domain cadherin repeat homology #label CR5\
686-717 #domain transmembrane #status predicted #label TMN\
718-877 #domain intracellular #status predicted #label INT\
836-849 #region serine-rich\
161,244,296,373, #binding_site carbohydrate (Asn) (covalent) #status
543,593,622,663 predicted
SUMMARY #length 877 #checksum 8685

Query Match 56.3%; Score 54; DB 1; Length 877;
Best Local Similarity 40.0%; Pred. No. 6.93e+00;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 271 LNSKIAFKIVSOEPA 285
Oy 1 LNSKIAFKIVSOEPA 15

RESULT 15
ENTRY IOHUC #type complete
```

```
TITLE cadherin 1 precursor - human
ALTERNATE_NAMES ARC-1; cell CAM 120/80; E-cadherin; epithelial cadherin;
ORGANISM L-CAM; uvomorulin
#formal_name Homo sapiens #common_name man
#formal_name Bos primigenius taurus #common_name cattle
#sequence_revision 30-Jun-1993 #text_change
26-Feb-1999
ACCESSIONS S37654; S31430; S05475; S31460; S06716; A57171; JC2230;
152294; 152704; S25141
REFERENCE S37654
#authors Busssemakers, M.J.G.; van Bokhoven, A.; Mees, S.G.M.; Kemler, R.; Schalken, J.A.
#journal Mol. Biol. Rep. (1993) 17:123-128
#title Molecular cloning and characterization of the human E-cadherin cDNA.
#accession S37654
##molecule_type mRNA
##residues 1-882 ##label BUS
##cross-references EMBL:Z13009; NID:g31072; PID:g31073
REFERENCE S31430
#authors Kelter, W.; Warda, A.; Oda, T.; Hirohashi, S.; Kemler, R.; Birchmeier, W.
#submission submitted to the EMBL Data Library, December 1992
#description Sequence of human E-cadherin cDNA.
#accession S31430
##molecule_type mRNA
##residues 1-542, 'F', 544-882 ##label KEL
##cross-references EMBL:Z18923; NID:g31074; PID:g31075
REFERENCE S05475
#authors Mansouri, A.; Spurr, N.; Goodfellow, P.N.; Kemler, R.
#journal Differentiation (1988) 38:67-71
#title Characterization and chromosomal localization of the gene encoding the human cell adhesion molecule uvomorulin.
#cross-references M01D:89031725
#accession S05475
##molecule_type mRNA
##residues 157-311 ##label MAN
##cross-references EMBL:X12790
#note nucleotide sequence is not complete
REFERENCE S31460
#authors Frixen, U.H.
#submission submitted to the EMBL Data Library, March 1990
#accession S31460
##molecule_type mRNA
##residues 265-392 ##label FRI
##cross-references EMBL:X52279; NID:g28821; PID:g28822
REFERENCE S06716
#authors Wheelock, M.J.; Buck, C.A.; Bechtol, K.B.; Damsky, C.H.
#journal J. Cell. Biochem. (1987) 34:187-202
#title Soluble 80-kd fragment of cell-CAM 120/80 disrupts cell-cell adhesion.
#cross-references M01D:87280410
#accession S06716
##molecule_type protein
#accession S06716
##residues 157, 157-162, 'V', 164-179 ##label WHE
REFERENCE A57171
#authors Bex, G.; Staes, K.; van Hengel, J.; Moelans, F.; Busssemakers, M.J.G.; van Bokhoven, A.; van Roy, F.
#journal Genomics (1995) 26:281-289
#title Cloning and characterization of the human invasion suppressor gene E-cadherin (CDH1).
#cross-references M01D:95324920
#accession A57171
##status nucleic acid sequence not shown; not compared with conceptual translation
##molecule_type DNA
##residues 1-30,33-882 ##label BER
##cross-references GB:J34784
REFERENCE JC2230
#authors Rimm, D.L.; Morrow, J.S.
#journal Biochem. Biophys. Res. Commun. (1994) 200:1754-1761
#title Molecular cloning of human E-cadherin suggests a novel subdivision of the cadherin superfamily.
#cross-references M01D:94242050
```

#accession JC2230
#molecule_type mRNA
#residues 1-9 'G', 11-15, 'RSPGSGERSPPCLRELHVHAPAPPEKRR', 52-67,
'1', 69, 'LPLP', 76-94, 'TDP', 98-99, 'GJR', 103-482, 'G',
484-529, 'R', 531-614, 'H', 616-633, 'RVP', 637-867, 'P',
869-881, 'H' ##label RIM

##cross-references GB:108599; NID:9340184; PID:9340185
#note the majority of differences between this and other
reports represent apparent frameshift errors

##note the authors translated the codon CCG for residue 868 as
Arg

REFERENCE
#authors Bussemakers, M.J.G.; Giroldi, L.A.; van Bokhoven, A.;
Schalken, J.A.
#journal Biochem. Biophys. Res. Commun. (1994) 203:1284-1290
#title Transcriptional regulation of the human E-cadherin gene in
human prostate cancer cell lines.
#cross-references MVID:94380041
#accession 152294
#status translation not shown; translated from GB/EMBL/DBJ

##molecule_type DNA
#residues 1-16 ##label RES
#cross-references GB:L34545; NID:9509604; PID:9509605
REFERENCE
#authors Becker, K.F.; Atkinson, M.J.; Reich, U.; Becker, I.; Nekarda,
H.; Siewert, J.R.; Hoefler, H.
#journal Cancer Res. (1994) 54:3845-3852
#title E-cadherin gene mutations provide clues to diffuse type
gastric carcinomas.
#cross-references MVID:94306394
#accession 152704
#status translated from GB/EMBL/DBJ

##molecule_type DNA
#residues 337-476 ##label RE2
#cross-references GB:S72492; NID:9632756

COMMENT Cadherins mediate calcium-dependent intercellular adhesion and are
thought to be involved in the sorting of different cell types
during morphogenesis.

GENETICS

#gene GDB:CDH1; UVO
#cross-references GDB:120484; OMTM:192090
#map_position 16q22.1-16q22.1
#introns 379/3; 440/3

CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;
transmembrane protein

FEATURE

1-27
28-154 #domain signal sequence #status predicted #label SIG\
155-882 #domain propeptide #status predicted #label PRO\
155-697 #product E-cadherin #status experimental #label MAT\
157-262 #domain extracellular #status predicted #label EXT\
232-237 #domain cadherin repeat homology #label CR1\
265-375 #region cadherin binding #status predicted
378-486 #domain cadherin repeat homology #label CR2\
487-595 #domain cadherin repeat homology #label CR3\
596-700 #domain cadherin repeat homology #label CR4\
698-731 #domain cadherin repeat homology #label CR5\
732-882 #domain transmembrane #status predicted #label TM\
840-853 #domain intracellular #status predicted #label INT\
#region serine-rich\
376,558,570,622,
637,849

SUMMARY #binding_site carbohydrate (Asn) (covalent) #status
predicted
#length 882 #molecular_weight 97455 #checksum 5112

Query Match 56.3%; Score 54; DB 1; Length 882;
Best local similarity 53.8%; Pred. NO. 6.93e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

DB 297 NAIATITLSQDP 309
I: | | | | | | | | | |
OY 2 NSKIAFRIVSQEP 14

Search completed: Fri Jun 11 17:26:59 1999
Job time : 88 secs.

KW CELL ADHESION; SIGNAL; TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
 KM CALCIUM-BINDING; REPEAT.
 FT SIGNAL 1 23
 FT PROPER 24 49
 FT CHAIN 50 999
 FT DOMAIN 50 615
 FT TRANSMEM 616 640
 FT TRANSMEM 641 999
 FT REPEAT 50 158
 FT REPEAT 159 268
 FT REPEAT 269 383
 FT REPEAT 384 499
 FT REPEAT 499 935
 FT REPEAT 935 966
 FT CARBOHYD 110 110
 FT CARBOHYD 180 180
 FT CARBOHYD 459 459
 FT CARBOHYD 545 545
 FT SEQUENCE 999 AA; 107503 MW; 4891FBAE CRC32;
 SQ

Query Match 100.0%; Score 96; DB 1; Length 999;
 Best Local Similarity 100.0%; Pred. No. 1,36e-11;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 190 LNSKIAFKIVSOEPA 204
 1 LNSKIAFKIVSOEPA 15

RESULT 2
 ID DSG1_BOVIN STANDARD; PRT; 1043 AA.
 AC 003763;
 DT 01-OCT-1993 (REL. 27, CREATED)
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN I) (DG1).
 GN DSG1.
 OS BOS TAURUS (BOVINE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINEAE; BOS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-MOZZLE EPITHELIUM;
 RA KOCH P.J., GOLDSCHMIDT M.D., ZIMBELMANN R., FRANK W.W.;
 RL SUBMITTED (MAR-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE OF 44-1043 FROM N.A.
 RC TISSUE-MOZZLE EPITHELIUM;
 RA KOCH P.J., WALSH M.J., SCHMELZ M., GOLDSCHMIDT M.D.,
 PA ZIMBELMANN R., FRANK W.W.;
 RT "Identification of desmoglein, a constitutive desmosomal
 glycoprotein, as a member of the cadherin family of cell adhesion
 molecules."
 RL EUR. J. CELL BIOL. 53:1-12(1990).
 RN [3]
 RP REVISIONS, AND SEQUENCE OF 101-123.
 RA MEDLINE; 92037656.
 RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R.,
 RA FRANK W.W.;
 RT "Complete amino acid sequence of the epidermal desmoglein precursor
 polypeptide and identification of a second type of desmoglein gene."
 RL EUR. J. CELL BIOL. 55:200-208(1991).
 RN [4]
 RP SEQUENCE OF 44-493 FROM N.A.
 RA MEDLINE; 91097553.
 RA GOODWIN L., HILL J.E., RAYNOR K., RASZI L., MANABE M., COMIN P.;
 RT "Desmoglein shows extensive homology to the cadherin family of cell
 adhesion molecules."
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 173:1224-1230(1990).
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.
 CC INVOLVED IN THE INTERACTION OF PLACQUE PROTEINS AND INTERMEDIATE
 CC FILAMENTS MEDIATING CELL-CELL ADHESION.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EPIDERMIS, MOZZLE, TONGUE AND OESOPHAGUS.
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE
 CC DESMOsome SUBFAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X58466; G307; -;
 DR EMBL; X57784; G436062; -;
 DR EMBL; M58155; G552318; -;
 DR FIR; S14603; IJB0GL.
 DR PROSITE; PS00232; CADHERIN; 2.
 DR PFAM; PF00028; cadherin; 3.
 DR HSSP; P09803; LEDH.
 DR CELL ADHESION; SIGNAL; TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
 KM CALCIUM-BINDING; REPEAT.
 FT SIGNAL 1 23
 FT PROPER 24 49
 FT CHAIN 50 1043
 FT DOMAIN 50 548
 FT TRANSMEM 549 573
 FT TRANSMEM 574 1043
 FT DOMAIN 50 158
 FT REPEAT 159 270
 FT REPEAT 271 385
 FT REPEAT 386 498
 FT REPEAT 499 845
 FT REPEAT 846 875
 FT REPEAT 876 905
 FT REPEAT 906 933
 FT REPEAT 934 962
 FT DOMAIN 963 1012
 FT CARBOHYD 110 110
 FT CARBOHYD 180 180
 FT CARBOHYD 496 496
 FT CONFLICT 124 124
 FT SEQUENCE 1043 AA; 112243 MW; 13898584 CRC32;
 SQ

Query Match 90.6%; Score 87; DB 1; Length 1043;
 Best Local Similarity 80.0%; Pred. No. 5.65e-09;
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 190 LNSKIAFKIVSOEPA 204
 1 LNSKIAFKIVSOEPA 15

RESULT 3
 ID DSG1_HUMAN STANDARD; PRT; 1049 AA.
 AC 002413;
 DT 01-OCT-1993 (REL. 27, CREATED)
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN I) (DG1).
 GN DSG1.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KERATINOCYTES;
 RA MEDLINE; 91271279.
 RA WHEELER G.N., PARKER A.E., THOMAS C.L., ATALIOFIS P., POYNTER D.,
 RA ARNEMAN J., RUTMAN A.J., PIDSLLEY S.C., WATT F.M., REES D.A.,
 RA BOXTON R.S., MAGEE A.I.;

RT "Desmosomal glycoprotein Dsg1, a component of intercellular desmosome junctions, is related to the cadherin family of cell adhesion molecules.";
RT PROC. NATL. ACAD. SCI. U.S.A. 88:4796-4800(1991).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSILLAR OESOPHAGUS.
CC -1- DOMAIN: CALCULIN MAY BE BOUND BY THE CADHERIN-LIKE REPEATS (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE DESMOSOMAL SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).
CC -----
CC EMBL; X56654; G30506; .
DR PIR; S16906; IJHUG1.
DR MIM; 125670; .
DR PROSITE; PS00232; CADHERIN; 2.
DR PFAM; PF00028; cadherin; 4.
DR HSSP; P09803; IEDH.
KW CELL ADHESION; SIGNAL; TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN; KW CALCULIN-BINDING; REPEAT.
FT SIGNAL 1 23
FT PROPEP 24 49
FT CHAIN 50 1049
FT DOMAIN 50 545
FT TRANSMEM 546 570
FT DOMAIN 571 1049
FT REPEAT 50 158
FT REPEAT 159 270
FT REPEAT 271 385
FT REPEAT 386 497
FT REPEAT 497 813
FT REPEAT 813 839
FT REPEAT 840 869
FT REPEAT 870 899
FT REPEAT 900 927
FT REPEAT 928 956
FT DOMAIN 969 1019
FT CARBOHYD 36 36
FT CARBOHYD 110 110
FT CARBOHYD 180 180
SQ SEQUENCE 1049 AA; 113715 MW; FDD79961 CRC32;

Query Match 90.6%; Score 87; DB 1; Length 1049;
Best Local Similarity 80.0%; Pred. No. 5,65e-09;
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 190 LNSKIAFKTRIOEPS 204
QY 1 LNSKIAFKTIVSOEPA 15
RESULT 4
ID DSG2 HUMAN STANDARD: PRT; 1117 AA.
AC 01A126;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DESMOGLEIN 2 PRECURSOR (HDGC).
GN DSG2.
OS HOMO SAPIENS (HUMAN).
OC EURAROTIA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.

RC TISSUE-COLON CARCINOMA;
RX MEDLINE; 94192736.
RA SCHAEFER S., KOCH P.J., FRANK W.W.;
RT "Identification of the ubiquitous human desmoglein, Dsg2, and the RT expression catalogue of the desmoglein subfamily of desmosomal RT cadherins.";
RL EXP. CELL RES. 211:391-399(1994).
RN [2]
RP SEQUENCE OF 777-1117 FROM N.A.
RX MEDLINE; 92037656.
RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMMERMANN R., FRANK W.W.;
RT "Complete amino acid sequence of the epidermal desmoglein precursor RT polypeptide and identification of a second type of desmoglein gene.";
RL EUR. J. CELL BIOL. 55:200-208(1991).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.
CC -1- DOMAIN: CALCULIN MAY BE BOUND BY THE CADHERIN-LIKE REPEATS (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE DESMOSOMAL SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).
CC -----
CC EMBL; Z26317; G416178; .
DR MIM; 125671; .
DR PROSITE; PS00232; CADHERIN; 3.
DR PFAM; PF00028; cadherin; 4.
DR HSSP; P15116; INCI.
KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL; KW CYTOSKELETON; CALCULIN-BINDING.
FT SIGNAL 1 23
FT PROPEP 24 48
FT CHAIN 49 1117
FT DOMAIN 49 608
FT TRANSMEM 609 633
FT DOMAIN 634 1117
FT REPEAT 49 159
FT REPEAT 160 272
FT REPEAT 273 387
FT REPEAT 388 502
FT REPEAT 502 880
FT REPEAT 880 911
FT REPEAT 912 941
FT REPEAT 942 967
FT REPEAT 968 991
FT REPEAT 992 1020
FT REPEAT 1021 1050
FT CARBOHYD 111 111
FT CARBOHYD 181 181
FT CARBOHYD 308 308
FT CARBOHYD 461 461
FT CARBOHYD 513 513
SQ SEQUENCE 1117 AA; 122385 MW; 84D3B898 CRC32;

Query Match 80.2%; Score 77; DB 1; Length 1117;
Best Local Similarity 73.3%; Pred. No. 3,44e-06;
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 191 LNSKISRYVSLPEA 205
QY 1 LNSKIAFKTIVSOEPA 15
RESULT 5
ID CAD1 MOUSE STANDARD: PRT; 884 AA.

AC P09803.061377;
 DT 01-MAR-1989 (REL. 10, CREATED)
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE EPIHELIAL-CADHERIN PRECURSOR (E-CADHERIN) (UWOMORULIN) (ARC-1).
 GN CDH1.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ICR;
 RA MEDLINE; 87315445.
 RA NAGAFUCHI A., SHIRAYOSHI Y., OKAZARI K., YASUDA K., TAKEICHI M.;
 RA "Transformation of cell adhesion properties by exogenously introduced
 RA E-cadherin cDNA.";
 PL NATURE 329:341-343(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-129/SV;
 RA MEDLINE; 92093614.
 RA RINGMALL M., BARIBAUD H., SCHMIDT C., KEMLER R.;
 RA "The structure of the gene coding for the mouse cell adhesion
 RA molecule uvomorulin.";
 PL NUCLEIC ACIDS RES. 19:6533-6539(1991).
 RN [3]
 RP SEQUENCE OF 174-884 FROM N.A., AND SEQUENCE OF 157-181.
 RC MEDLINE; 8811553.
 RA RINGMALL M., SCHUH R., VESTWEBER D., EISTETER H., LOTTSPRECH F.,
 RA ENGEL J., DOELZ R., JAEHNIG F., EPPLER J., MAYR S., MUELLER C.,
 RA KEMLER R.;
 RA "The structure of cell adhesion molecule uvomorulin. Insights into
 RA the molecular mechanism of Ca²⁺-dependent cell adhesion.";
 PL EMBO J. 6:3647-3653(1987).
 RN [4]
 RP SEQUENCE OF 1-15 FROM N.A.
 RC MEDLINE; 92107977.
 RA BEHRENS J., LOEWENICK O., KLEIN-HITPASS L., BIRCHMEIER W.;
 RA "The E-cadherin promoter: functional analysis of a G-C-rich region
 RA and an epithelial cell-specific palindromic regulatory element.";
 PL PROC. NATL. ACAD. SCI. U.S.A. 88:11495-11499(1991).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 157-370.
 RC MEDLINE; 96176249.
 RA NAGAR B., OVERDUIN M., IKURA M., RINT J.M.;
 RA "Structural basis of calcium-induced E-cadherin rigidification and
 RA dimerization.";
 PL NATURE 380:360-364(1996).
 RN [6]
 RP STRUCTURE BY NMR OF 157-260.
 RC MEDLINE; 96271285.
 RA OVERDUIN M., TONG K.I., KAY C.M., IKURA M.;
 RA "1H, 15N and 13C resonance assignments and monomeric structure of the
 RA amino-terminal extracellular domain of epithelial cadherin.";
 PL J. BIOMOL. NMR 7:173-189(1996).
 RN [7]
 RP FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
 CC SORTING OF HETEROGENEOUS CELL TYPES.
 CC [8]
 CC SUBUNIT: HOMODIMER.
 CC [9]
 CC TISSUE SPECIFICITY: NON-NEURAL EPITHELIAL TISSUES.
 CC [10]
 CC SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
 CC [11]
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; X06115; G50765; -

DR EMBL; X06961; G818011; -
 DR EMBL; X06962; G818011; JOINED.
 DR EMBL; X06963; G818011; JOINED.
 DR EMBL; X06964; G818011; JOINED.
 DR EMBL; X06965; G818011; JOINED.
 DR EMBL; X06966; G818011; JOINED.
 DR EMBL; X06967; G818011; JOINED.
 DR EMBL; X06968; G818011; JOINED.
 DR EMBL; X06969; G818011; JOINED.
 DR EMBL; X06970; G818011; JOINED.
 DR EMBL; X06971; G818011; JOINED.
 DR EMBL; X06972; G818011; JOINED.
 DR EMBL; X06973; G818011; JOINED.
 DR EMBL; X06974; G818011; JOINED.
 DR EMBL; X06975; G818011; JOINED.
 DR EMBL; X06339; G55151; -
 DR EMBL; M81449; G192326; -
 DR PIR; S04528; I0MSCE.
 DR PDB; 1EDH; 11-JAN-97.
 DR PDB; 1SUH; 11-JUL-96.
 DR MGD; MGI:88354; CDH1.
 DR PROSITE; PS00232; CADHERIN; 3.
 DR PIR; PF01049; Cadherin_C-term; 1.
 DR CELL ADHESION; GLYCOPROTEIN; PHOSPHORYLATION; TRANSMEMBRANE;
 DR SIGNAL; 27
 DR PROPEP; 28
 DR CHAIN; 157
 DR DOMAIN; 157
 DR TRANSMEM; 709
 DR DOMAIN; 733
 DR DOMAIN; 734
 DR REPEAT; 157
 DR REPEAT; 265
 DR REPEAT; 378
 DR REPEAT; 488
 DR REPEAT; 489
 DR REPEAT; 595
 DR REPEAT; 596
 DR DOMAIN; 840
 DR CARBOHYD; 560
 DR CARBOHYD; 639
 DR CONFLICT; 267
 DR CONFLICT; 272
 DR SEQUENCE; 884 AA; 98255 MW; 8C758910 CRC32;
 SO
 Query Match 62.5%; Score 60; DB 1; Length 884;
 Best Local Similarity 61.5%; Pred. No. 7.44e-02;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 Db 299 NAAIATVTSODP 311
 Oy 2 NSKIAFKIVSOEP 14
 ID CAD2_HUMAN STANDARD; PRT; 906 AA.
 AC P19022; Q14923;
 DT 01-NOV-1990 (REL. 16, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE HUMAN N-CADHERIN; NUCLEOTIDE AND DEDUCED AMINO ACID SEQUENCE.";
 DE CDH2 OR CDH2N OR NCAD.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE; 91016946.
 RA REID R.A., HEMPERLY J.J.;
 RA "Human N-cadherin: nucleotide and deduced amino acid sequence.";
 PL NUCLEIC ACIDS RES. 18:5896-5896(1990).
 RN [2]
 RP REVISIONS TO 341; 699 AND 705.
 RA REID R.A.;

RL SUBMITTED (NOV-1990) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92363956.
 RA SALOMON D., AVALON O., PATEL-KING R., HYNES R.O., GEIGER B.;
 RT "Extrajunctional distribution of N-cadherin in cultured human
 RT endothelial cells."
 RT J. CELL SCI. 102:7-17(1992).
 RN [4]
 RP SEQUENCE OF 160-906 FROM N.A.
 RX MEDLINE: 90347462.
 RA WALSH F.S., BARTON C.H., PUTT W., MOORE S.E., KELSELL D.,
 RA SPURR N., GOODFELLOW P.N.;
 RT "N-cadherin gene maps to human chromosome 18 and is not linked to the
 RT E-cadherin gene."
 RT J. NEUROCHEM. 55:805-812(1990).
 RN [5]
 RP SEQUENCE OF 1-20 FROM N.A.
 RX MEDLINE: 95048366.
 RA WALLIS J.A., FOX M., WALSH F.S.;
 RT "Structure of the human N-cadherin gene: YAC analysis and fine
 RT chromosomal mapping to 18q11.2."
 RT GENOMICS 22:172-179(1994).
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
 CC SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN
 CC NEURONAL RECOGNITION MECHANISM.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC -----
 DR EMBL: X57548; E31840; -
 DR EMBL: X54315; G34999; -
 DR EMBL: S42303; G253483; -
 NR EMBL: M34064; G416293; -
 DR EMBL: Z27420; G806346; -
 DR PIR: A38870; IJHUCN.
 DR MIM: 114020; -
 DR PROSITE: PS00232; CADHERIN; 3.
 DR PFAM: PF00028; cadherin; 5.
 DR PFAM: PF01049; Cadherin_C-term; 1.
 DR HSSP: P15116; INCH.
 KW CELL ADHESION; GLYCOPROTEIN; PHOSPHORYLATION; TRANSMEMBRANE;
 KW CALCIUM-BINDING; REPEAT; SIGNAL.
 FT SIGNAL 1 23
 FT PROPEP 24 159
 FT CHAIN 160 906
 FT DOMAIN 160 724
 FT TRANSMEM 725 746
 FT DOMAIN 747 906
 FT REPEAT 160 267
 FT REPEAT 268 382
 FT REPEAT 383 487
 FT REPEAT 488 603
 FT REPEAT 604 714
 FT DOMAIN 863 878
 FT CARBOHYD 190 190
 FT CARBOHYD 273 273
 FT CARBOHYD 325 325
 FT CARBOHYD 402 402
 FT CARBOHYD 572 572
 FT CARBOHYD 651 651
 FT CARBOHYD 692 692
 FT CONFLICT 12 12
 FT CONFLICT 16 16
 L -> A (IN REF. 3 AND 5).

FT CONFLICT 196 196 S -> T (IN REF. 1).
 FT CONFLICT 212 212 I -> L (IN REF. 4).
 FT CONFLICT 357 357 N -> I (IN REF. 1).
 FT CONFLICT 867 867 A -> L (IN REF. 3).
 SQ SEQUENCE 906 AA; 99851 MW; BBLF9558 CRC32;
 Query Match 62.5%; Score 60; DB 1; Length 906;
 Best Local Similarity 46.7%; Pred. No. 7.44e-02;
 Matches 7; Conservative 6; Mismatches 2; Indels 0; Gaps 0;
 Db 300 LNCMLRXYRVSQAPS 314
 1 LNSKIAFKIVSOEPA 15
 QY
 RESULT 7
 ID YOX1_YEAST STANDARD; PRT; 385 AA.
 AC P34161;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE HOMEOBOX PROTEIN YOX1.
 GN YOX1 OR YML027W.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCETES; SACCHAROMYCETALES;
 OC SACCCHAROMYCETACEAE; SACCHAROMYCES.
 RN [1]
 RP SEQUENCE OF 1-296 FROM N.A.
 RX MEDLINE: 93209080.
 RA KAUFMANN E.;
 RT "In vitro binding to the leucine tRNA gene identifies a novel yeast
 RT homeobox gene."
 RT N. N. CHROMOSOMA 102:174-179(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-9288C / A9972;
 RA BADCOCK K., CHURCHER C., BARRELL B.G., RAJANDREAN M.A., WALSH S.V.;
 RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- FUNCTION: IN VITRO, IS CAPABLE OF BINDING TO THE DNA OF THE
 CC LEUCINE TRNA GENE.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: WITH OTHER HOMEOBOX PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC -----
 DR EMBL: X62392; G5502; -
 DR EMBL: Z46659; G575692; -
 DR PIR: S33388; S33388.
 DR SGD: L0002540; YOX1.
 DR PROSITE: PS00027; HOMEOBOX_1; 1.
 DR PROSITE: PS50071; HOMEOBOX_2; 1.
 DR PFAM: PF00046; homeobox; 1.
 DR HSSP: P06601; LFYL.
 KW HOMEOBOX; DNA-BINDING; NUCLEAR PROTEIN.
 FT DNA_BIND 176 235
 FT CONFLICT 291 296
 FT CONFLICT 385 AA; 42739 MW; IESBBA9 CRC32;
 SQ
 Query Match 61.5%; Score 59; DB 1; Length 385;
 Best Local Similarity 63.6%; Pred. No. 1.28e-01;
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 Db 364 KFGFKIVDOOP 374
 4 KIFKIVSOEP 14
 QY


```

FT DISULFID 4080 4089 BY SIMILARITY.
FT DISULFID 4096 4107 BY SIMILARITY.
FT DISULFID 4101 4116 BY SIMILARITY.
FT DISULFID 4118 4127 BY SIMILARITY.
FT DISULFID 4325 4341 BY SIMILARITY.
FT DISULFID 4334 4350 BY SIMILARITY.
FT DISULFID 4352 4361 BY SIMILARITY.
FT DISULFID 239 239 POTENTIAL.
FT CARBOHYD 257 257 POTENTIAL.
FT CARBOHYD 276 276 POTENTIAL.
FT CARBOHYD 280 280 POTENTIAL.
FT CARBOHYD 402 402 POTENTIAL.
FT CARBOHYD 461 461 POTENTIAL.
FT CARBOHYD 605 605 POTENTIAL.
FT CARBOHYD 631 631 POTENTIAL.
FT CARBOHYD 1155 1155 POTENTIAL.
FT CARBOHYD 1367 1367 POTENTIAL.
FT CARBOHYD 1458 1458 POTENTIAL.
FT CARBOHYD 1751 1751 POTENTIAL.
FT CARBOHYD 1831 1831 POTENTIAL.
FT CARBOHYD 1880 1880 POTENTIAL.
FT CARBOHYD 2080 2080 POTENTIAL.
FT CARBOHYD 2171 2171 POTENTIAL.
FT CARBOHYD 2247 2247 POTENTIAL.
FT CARBOHYD 2290 2290 POTENTIAL.
FT CARBOHYD 2437 2437 POTENTIAL.
FT CARBOHYD 2581 2581 POTENTIAL.
FT CARBOHYD 2799 2799 POTENTIAL.
FT CARBOHYD 2920 2920 POTENTIAL.
FT CARBOHYD 2946 2946 POTENTIAL.
FT CARBOHYD 2967 2967 POTENTIAL.
FT CARBOHYD 3167 3167 POTENTIAL.
FT CARBOHYD 3303 3303 POTENTIAL.
FT CARBOHYD 3385 3385 POTENTIAL.
FT CARBOHYD 3389 3389 POTENTIAL.
FT CARBOHYD 3525 3525 POTENTIAL.
FT CARBOHYD 3852 3852 POTENTIAL.
FT CARBOHYD 3865 3865 POTENTIAL.
FT CARBOHYD 3905 3905 POTENTIAL.
FT CARBOHYD 4306 4306 POTENTIAL.
FT CARBOHYD 4414 4414 POTENTIAL.
FT CARBOHYD 4471 4471 POTENTIAL.
FT CARBOHYD 4487 4487 POTENTIAL.
FT CARBOHYD 4539 4539 POTENTIAL.
FT CARBOHYD 4550 4550 POTENTIAL.
FT CARBOHYD 1229 1229 POTENTIAL.
FT VARIANT 1233 1233 S -> G.
SQ SEQUENCE 5147 AA: 564868 MW: 1EF20E13 CRC32;

```

```

Query Match 59.4%; Score 57; DB 1; Length 5147;
Best Local Similarity 50.0%; Pred. No. 3,70e-01;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```

```

Db 1520 LNSKIAFKIVSQEP 1533
OY 1 LNSKIAFKIVSQEP 14

```

```

RESULT 10
ID EFP-BORBU STANDARD; PRT; 192 AA.
AC 051232;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ELONGATION FACTOR P (EF-P).
GN EFP OR B0214.
OS BORRELLIA BURDOFFERI (LYME DISEASE SPIROCHETE).
OC BORRELLIA SPIROCHAETALES; SPIROCHAETACEAE; BORRELLIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 35210 / B31;
RX MDLINE; 98065943.
RA FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,

```

```

RA LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
RA PETERSON J., KERLAVAGE A.R., OUCKENBUSH J., SALZBERG S., HANSON M.,
RA VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
RA UTTERBACK T., MATTHEY L., McDONALD L., ARTIACH P., BOWMAN C.,
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
RA SMITH H.O., VENTER J.C.;
RA "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi."
RT NATURE 390:580-586(1997).
RL - FUNCTION: INVOLVED IN PEPTIDE BOND SYNTHESIS. STIMULATE EFFICIENT
CC TRANSLATION AND PEPTIDE-BOND SYNTHESIS ON NATIVE OR RECONSTITUTED
CC 70S RIBOSOMES IN VITRO. PROBABLY FUNCTIONS INDIRECTLY BY ALTERING
CC THE AFFINITY OF THE RIBOSOME FOR AMINOACYL-TRNA, THUS INCREASING
CC THEIR REACTIVITY AS ACCEPTORS FOR PEPTIDYL TRANSFERASE
CC (BY SIMILARITY).
CC -1- PATHWAY: PROTEIN BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE ELONGATION FACTOR P FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001132; G2688116; -
DR TIGR; B0214; -
DR PROSITE; PS01275; EFP, FALSE_NEG.
DR PROTEIN BIOSYNTHESIS; ELONGATION FACTOR.
SQ SEQUENCE 192 AA: 21413 MW: CB74964B CRC32;

```

```

Query Match 57.3%; Score 55; DB 1; Length 192;
Best Local Similarity 53.3%; Pred. No. 1,04e+00;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

Db 130 LAPRIAEVEVEENA 144
OY 1 LNSKIAFKIVSQEP 15

```

```

RESULT 11
ID CAD2_BOVIN STANDARD; PRT; 877 AA.
AC P19534;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE NEURAL-CADHERIN PRECURSOR (N-CADHERIN) (FRAGMENT).
GN CHD2.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINA; BOS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90360979.
RA LIAM C.W., CANNON C., POWER M.D., KIBONEXA P.K., RUBIN L.L.;
RT "Identification and cloning of two species of cadherins in bovine
RT endothelial cells."
RL EMO J. 9:2701-2708(1990).
CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS. CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN
CC NEURONAL RECOGNITION MECHANISM.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC -----

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: X53615; G664894; .

DR PIR: S11693; IJBOCN.

DR PROSITE: PS00232; CADHERIN; 3.

DR PFAM: PF00028; cadherin; 5.

DR PFAM: PF01049; Cadherin_C-term; 1.

DR HSSP: P51116; INCH.

KW CELL ADHESION; GLYCOPROTEIN; PHOSPHORYLATION; TRANSMEMBRANE;

KM CALCULUM-BINDING; REPEAT.

FT NON_TER 1

FT PROPEP <1 130

FT CHAIN 131 877 N-CADHERIN.

FT DOMAIN 131 695 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 696 717 POTENTIAL.

FT DOMAIN 718 877 CYTOPLASMIC (POTENTIAL).

FT REPEAT 131 238 CADHERIN 1.

FT REPEAT 239 363 CADHERIN 2.

FT REPEAT 364 468 CADHERIN 3.

FT REPEAT 469 574 CADHERIN 4.

FT REPEAT 575 685 CADHERIN 5.

FT DOMAIN 834 849 SER-RICH.

FT CARBOHYD 161 161 POTENTIAL.

FT CARBOHYD 244 244 POTENTIAL.

FT CARBOHYD 296 296 POTENTIAL.

FT CARBOHYD 373 373 POTENTIAL.

FT CARBOHYD 543 543 POTENTIAL.

FT CARBOHYD 622 622 POTENTIAL.

FT CARBOHYD 663 663 POTENTIAL.

SQ SEQUENCE 877 AA; 96845 MW; 8E417A39 CRC32;

Query Match 56.3%; Score 54; DB 1; Length 877;

Best Local Similarity 40.0%; Pred. No. 1.74e+00;

Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 271 LNSKIRYISQAPS 285

1 LNSKIRYISQAPS 15

RESULT 12

ID CAD1_HUMAN STANDARD; PRT; 882 AA.

AC P12830; Q14216; .

DT 01-OCT-1989 (REL. 12, CREATED)

DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE EPIHELIAL-CADHERIN PRECURSOR (E-CADHERIN) (VIMOMORULIN) (CAM 120/80).

GN CDH1 OR UVO OR CDHE.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC PRIMATES; CATHARTINI; HOMINIDAE; HOMO.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 93111394.

RA BUSSEMAKERS M.J.G., MEES S.G.M., VAN BOKHOVEN A., DEBRUYNE F.M.J., SCHALKEN J.A.;

RT "Molecular cloning and characterization of the human E-cadherin cDNA.";

RL WOL. BIOL. REP. 17:123-128(1993).

RN [2]

RP SEQUENCE FROM N.A.

RA KELLER W., WARD A., ODA T., HIROHASHI S., KELLER R., BIRCHMEIER W.;

RL SUBMITTED (XXX-1992) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [3]

RP SEQUENCE OF 172-311 FROM N.A.

RC TISSUE-LIVER.

RX MEDLINE; 89031725.

RA MANSOORI A., SPURR N., GOODFELLOW P.N., KELLER R.;

RT "Characterization and chromosomal localization of the gene encoding the human cell adhesion molecule uvomorulin.";

RL DIFFERENTIATION 38:67-71(1988).

RN [4]

RP SEQUENCE OF 1-16 FROM N.A.

RX MEDLINE; 94380041.

RA BUSSEMAKERS M.J., GIROLDI L.A., VAN BOKHOVEN A., SCHALKEN J.A.;

RT "Transcriptional regulation of the human E-cadherin gene in human prostate cancer cell lines: characterization of the human E-cadherin gene promoter.";

RL BIOCHEM. BIOPHYS. RES. COMMUN. 203:1284-1290(1994).

RN [5]

RP REVIEW ON VARIANTS.

RX MEDLINE; 98415721.

RA BERR G., BECKER K.-F., HOEFELER H., VAN ROY F.;

RT "Mutations of the human E-cadherin (CDH1) gene.";

RL HUM. MUTAT. 12:226-237(1998).

RN [6]

RP VARIANTS ALA-370 AND ASN-473.

RX MEDLINE; 94306394.

RA BECKER K.-F., ATKINSON M.J., REICH U., BECKER I., NEKARDA H., SIEWERT J.R., HOEFELER H.;

RT "E-cadherin gene mutations provide clues to diffuse type gastric carcinomas.";

RL CANCER RES. 54:3845-3852(1994).

RN [7]

RP VARIANTS THR-617, VAL-711 AND GLY-838.

RX MEDLINE; 94355985.

RA RISINGER J.I., BERCHECK A., KOHLER M.F., BOYD J.;

RT "Mutations of the E-cadherin gene in human gynecologic cancers.";

RL NAT. GENET. 7:98-102(1994).

CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS. THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE SORTING OF HETEROGENEOUS CELL TYPES. E-CADHERIN HAS A POTENT INVASIVE SUPPRESSOR ROLE.

CC -1- SUBUNIT: HOMODIMER.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- TISSUE SPECIFICITY: NON-NEURAL EPITHELIAL TISSUES.

CC -1- DEFECTS OF CDH1 IS REGARDED AS ONE OF THE MAIN MOLECULAR EVENTS INVOLVED IN DYSFUNCTION OF THE CELL-CELL ADHESION SYSTEM, TRIGGERING CANCER INVASION (GASTRIC, BREAST, OVARY, ENDOMETRIUM AND THYROID) AND METASTASIS.

CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: Z13009; G31073; .

DR EMBL: Z18923; G31075; .

DR EMBL: X12790; G930046; .

DR EMBL: L34345; G509605; .

DR PIR: S25141; IJHUCE.

DR PIR: S37654; S37654.

DR MIM: 192090; .

DR PROSITE: PS00232; CADHERIN; 3.

DR PFAM: PF00028; cadherin; 5.

DR HSSP: P09803; ISUH.

DR PFAM: PF01049; Cadherin_C-term; 1.

KW CELL ADHESION; GLYCOPROTEIN; PHOSPHORYLATION; TRANSMEMBRANE;

KM CALCULUM-BINDING; REPEAT; SIGNAL; DISEASE MUTATION.

FT SIGNAL 1 27

FT PROPEP 28 154

FT CHAIN 155 882

FT DOMAIN 155 707

FT TRANSMEM 708 731

FT DOMAIN 732 882

FT REPEAT 155 262

FT REPEAT 263 375

FT REPEAT 376 486

FT REPEAT 487 593

FT REPEAT 594 697

EXTRACELLULAR (POTENTIAL).

POTENTIAL.

CYTOPLASMIC (POTENTIAL).

CADHERIN 1.

CADHERIN 2.

CADHERIN 3.

CADHERIN 4.

CADHERIN 5.

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M31131; G309125; -
 DR EMBL: AB008811; D1024428; -
 DR PIR: A32759; IJMSCN.
 DR PDB: INGG; 10-JUL-95.
 DR PDB: INCI; 10-JUL-95.
 DR MGD: MGI:88355; CDH2.
 DR PROSITE: PS00232; CADHERIN; 3.
 DR PFAM: PF00028; cadherin; 5.
 DR PFAM: PF01049; Cadherin_C-term; 1.
 KW CELL ADHESION; GLYCOPROTEIN; PHOSPHORYLATION; TRANSMEMBRANE;
 KM CALCULUM-BINDING; REPEAT; SIGNAL; 3D-STRUCTURE.
 FT SIGNAL 1 23
 FT PROPEP 24 159
 FT CHAIN 160 906
 FT DOMAIN 160 724
 FT TRANSMEM 725 746
 FT DOMAIN 747 906
 FT REPEAT 160 267
 FT REPEAT 268 382
 FT REPEAT 383 497
 FT REPEAT 498 603
 FT REPEAT 604 717
 FT DOMAIN 863 878
 FT CARBOHYD 190 190
 FT CARBOHYD 273 273
 FT CARBOHYD 323 323
 FT CARBOHYD 402 402
 FT CARBOHYD 572 572
 FT CARBOHYD 651 651
 FT CARBOHYD 692 692
 FT CONFLICT 7 9
 FT CONFLICT 565 565
 FT CONFLICT 567 567
 FT CONFLICT 624 624
 SQ SEQUENCE 906 AA; 99761 MW; 4C13AB40 CRC32;

Query Match 56.3%; Score 54; DB 1; Length 906;
 Best Local Similarity 40.0%; Pred. No. 1.74e+00;
 Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 300 LINGULARITYSOAPS 314
 |||||
 QY 1 LNSKIATKIVSOEPA 15

RESULT 15
 ID AS10-YEAST STANDARD; PRT; 1146 AA.
 AC P48361;
 DT 01-FEB-1996 (REL. 33, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE ASK10 PROTEIN.
 DE ASK10 OR YGR097W.
 CC SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 CC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
 CC SACCHAROMYCETACEAE; SACCHAROMYCETES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / SEY6210;
 KX MEDLINE; 97060018.
 TA PAGE N. SHERATON J., BROWN J.L., STEWART R.S., BUSSEY H.;
 TI "Identification of Ask10 as a multicopy activator of Skn7p-dependent
 transcription of a His3 reporter gene";
 J. YEAST 12:267-272 (1996).
 RT [2]

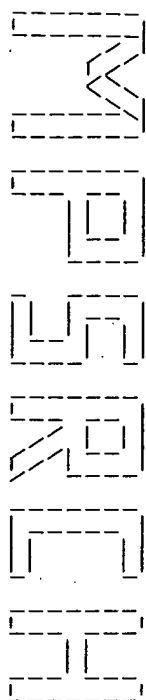
RP SEQUENCE FROM N.A.
 RA HERANDEZ K., WEBER N., WIPPLI P., SCHMIDHEINI T.;
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- FUNCTION: PUTATIVE ACTIVATOR OF SKN7.
 CC -1- SIMILARITY: TO YEAST YIL105C AND YNL047C.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U27209; G829640; -
 DR EMBL: Z72882; E243478; -
 DR SGD: L0002770; ASK10.
 FT DOMAIN 22 26
 FT DOMAIN 625 628
 FT DOMAIN 933 938
 FT DOMAIN 958 961
 FT DOMAIN 972 975
 FT CONFLICT 57 57
 FT CONFLICT 346 346
 FT CONFLICT 464 464
 FT CONFLICT 467 467
 FT CONFLICT 603 603
 FT CONFLICT 906 906
 SQ SEQUENCE 1146 AA; 126863 MW; 72D6FA87 CRC32;

Query Match 56.3%; Score 54; DB 1; Length 1146;
 Best Local Similarity 70.0%; Pred. No. 1.74e+00;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 692 TERIVSPEPS 701
 |||||
 QY 6 AFKIVSOEPA 15

Search completed: Fri Jun 11 17:22:53 1999
 Job time : 8 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_p protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 11 17:23:10 1999; MasPar time 6.15 Seconds

Tabular output not generated. 133.169 Million cell updates/sec

Title: >US-08-991-628-3
Description: (1-15) from US08991628.dep
Perfect Score: 96
Sequence: 1 LNSKIAFKIVSQEPA 15

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

sptrembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 25.449; Variance 30.607; scale 0.831

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	92	95.8	993	11	035902	3.46e-09
2	61	63.5	1421	10	080907	1.65e-01
3	58	60.4	138	2	P72670	7.46e-01
4	55	57.3	225	5	Q20152	3.21e+00
5	55	57.3	327	5	Q93345	3.21e+00
6	54	56.3	108	10	048607	5.16e+00
7	54	56.3	238	11	055075	5.16e+00
8	54	56.3	878	4	015855	5.16e+00
9	54	56.3	2258	14	089238	5.16e+00
10	54	56.3	2258	14	081081	5.16e+00
11	53	55.2	464	4	043159	8.24e+00
12	53	55.2	775	10	P93205	8.24e+00
13	53	55.2	1980	10	084185	8.24e+00
14	53	55.2	2228	14	055528	8.24e+00
15	53	55.2	2228	14	055530	8.24e+00
16	53	55.2	2228	14	098705	8.24e+00
17	53	55.2	2910	5	Q26008	8.24e+00
18	52	54.2	248	5	Q20221	1.31e+01
19	52	54.2	306	5	Q20776	1.31e+01
20	52	54.2	485	2	083862	1.31e+01

21	52	54.2	697	5	076208	PROPHENOLOXIDASE.	1.31e+01
22	52	54.2	709	11	088977	CALPAIN LP82.	1.31e+01
23	52	54.2	709	11	008702	CALPAIN LP82.	1.31e+01
24	52	54.2	737	11	070376	CALPAIN ISOFORM LP85.	1.31e+01
25	52	54.2	763	11	070482	MUSCLE TYPE CALPAIN P9	1.31e+01
26	52	54.2	821	6	046596	SKELETAL MUSCLE SPECIF	1.31e+01
27	52	54.2	914	11	088199	CADHERIN-RELATED NEURA	1.31e+01
28	51	53.1	103	2	049268	HYPOTHETICAL PROTEIN (2.06e+01
29	51	53.1	216	5	P91446	COSMID T16A1.	2.06e+01
30	51	53.1	241	8	Q33766	RIBOSOMAL PROTEIN S3.	2.06e+01
31	51	53.1	813	4	075229	R31449.3 (FRAGMENT).	2.06e+01
32	51	53.1	2233	14	081890	L PROTEIN.	2.06e+01
33	51	53.1	2764	5	001398	NEUROFIBROMIN.	2.06e+01
34	51	53.1	2802	5	001398	NEUROFIBROMIN.	2.06e+01
35	51	53.1	2802	5	001397	BETA 8 INTEGRIN (FRAGM	3.23e+01
36	50	52.1	194	13	091415	T04C10.4 PROTEIN.	3.23e+01
37	50	52.1	208	5	022156	PROTEINASE INHIBITOR.	3.23e+01
38	50	52.1	221	10	041482	PROTEINASE INHIBITOR.	3.23e+01
39	50	52.1	221	10	041481	YDB PROTEIN.	3.23e+01
40	50	52.1	261	2	P96679	NON-INDUCIBLE CARBOXY	3.23e+01
41	50	52.1	276	11	008559	INDUCIBLE CARBOXYL RED	3.23e+01
42	50	52.1	277	11	008558	COSMID W02D3.	3.23e+01
43	50	52.1	355	5	001810	T13D8.29 PROTEIN.	3.23e+01
44	50	52.1	578	10	080763	ORFL12.	3.23e+01
45	50	52.1	586	9	037867		3.23e+01

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	993 AA.
ID	035902			
AC	035902;			
DT	01-JAN-1998 (TREMBLREL. 05, CREATED)			
DT	01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)			
DE	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	DESMOGLLEIN 3 (FRAGMENT).			
GN	DSG3.			
OS	MUS MUSCULUS (MOUSE).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;			
OC	SCURIONATHI; MURIDAE; MURINAE; MUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BALB/C;			
RA	ISHIKAWA H., LI K., UTTIO J.;			
RL	SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.			
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).			
DR	EMBL: U86016; G2290200; -			
DR	PROSITE: PS00232; CADHERIN; 2.			
DR	PRAM; PF00028; cadherin; 4.			
KW	CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.			
FT	NON_TER 993			
SO	SEQUENCE 993 AA; 10788 MW; 881794BD CRC32;			
Query Match	95.88;	Score 92;	DB 11;	Length 993;
Best Local Similarity	93.38;	Pred. No. 3.46e-09;		
Matches	14;	Conservative	1;	Mismatches 0; Indels 0; Gaps 0;
Db	190 LNSKIAFKIVSQEPA 204			
Qy	1 LNSKIAFKIVSQEPA 15			
RESULT	2	PRELIMINARY;	PRT;	1421 AA.
ID	080907			
AC	080907;			
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)			
DE	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	T19C21.7 PROTEIN.			
GN	T19C21.7.			
OS	ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).			
OC	EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;			
OC	EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;			

CC CAPPARALE; BRASSICACEAE; ARABIDOPSIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV, COLUMBIA;
 RA ROUNSELEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
 RA STYES S.M., KAUL S., MASON T.M., KERLAJAGE A.R., ADAMS M.D.,
 RA SOMERVILLE C.R., VENTER J.C.,
 RT "Arabidopsis thaliana chromosome II BAC T19C21 genomic sequence."
 RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AC004683; G3395428;
 SQ SEQUENCE 1421 AA; 154325 MW; 1B6D2FF1 CRC32;
 Query Match 63.5%; Score 61; DB 10; Length 1421;
 Best Local Similarity 40.0%; Pred. No. 1.65e+01;
 Matches 6; Conservative 8; Mismatches 1; Indels 0; Gaps 0;
 Db 262 VDSKSYEITONPG 276
 QY 1 LNSKAFKIVSOEPA 15
 RESULT 3
 ID P72670 PRELIMINARY; PRT; 138 AA.
 AC P72670.
 DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
 DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DC HYPOTHEICAL 16.6 KD PROTEIN.
 CS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
 CC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC6803;
 RA TABATA S.,
 RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC6803;
 RC MEDLINE: 97061201.
 RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
 RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAKI N., KIMURA T.,
 RA HOSUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARDO K., OKUMURA S.,
 RA SHIMO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M.,
 RA IABATA S.,
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions."
 RL DNA RES. 3:109-136(1996).
 DR EMBL: D90899; D1017405;
 KW HYPOTHEICAL PROTEIN.
 SQ SEQUENCE 138 AA; 16580 MW; CA436352 CRC32;
 Query Match 60.4%; Score 58; DB 2; Length 138;
 Best Local Similarity 38.5%; Pred. No. 7.46e+01;
 Matches 5; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
 Db 28 AALAEVLPDPT 40
 QY 3 SKIAFKIVSOEPA 15
 RESULT 4
 ID Q20152 PRELIMINARY; PRT; 225 AA.
 AC Q20152;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE F3887.4 PROTEIN.
 GN F3887.4.
 OS CAENORHABDITIS ELEGANS.
 CC EUKARYOTA; METAZOA; NEMATODA; SECERNTEA; RHABDITIA; RHABDITIDA;
 CC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 RN [1]

RP SEQUENCE FROM N.A.
 RA LENNARD N.,
 RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COOPER T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SPALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans."
 RL NATURE 368:32-38(1994).
 DR EMBL: Z74033; E1346602;
 SQ SEQUENCE 225 AA; 26356 MW; 66EBA455 CRC32;
 Query Match 57.3%; Score 55; DB 5; Length 225;
 Best Local Similarity 53.3%; Pred. No. 3.21e+00;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 Db 205 LASKYFVRSQVPS 219
 QY 1 LNSKAFKIVSOEPA 15
 RESULT 5
 ID Q93345 PRELIMINARY; PRT; 327 AA.
 AC Q93345;
 DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE C36B1.11 PROTEIN.
 GN C36B1.11.
 OS CAENORHABDITIS ELEGANS.
 CC EUKARYOTA; METAZOA; NEMATODA; SECERNTEA; RHABDITIA; RHABDITIDA;
 CC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA LENNARD N.,
 RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COOPER T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SPALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans."
 RL NATURE 368:32-38(1994).
 DR EMBL: Z80215; E1344392;
 SQ SEQUENCE 327 AA; 37139 MW; DBA2C259 CRC32;
 Query Match 57.3%; Score 55; DB 5; Length 327;
 Best Local Similarity 80.0%; Pred. No. 3.21e+00;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 81 KLAFKIVSOE.90
 QY 4 KIAFKIVSOE 13

```

RESULT 6
ID 048607 PRELIMINARY; PRT; 108 AA.
AC 048607;
DT 01-JUN-1998 (TREMREL. 06, CREATED)
DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 12.3 KD PROTEIN (FRAGMENT).
OS HORDEUM VULGARE (BARLEY).
OC EUKARYOTA: VIRIDIPHYTES: STREPTOPHYTA: EMBRYOPHYTA: TRACHEOPHYTA:
OC EUPHYLOPHYTES: SPERMATOPHYTES: MAGNOLIOPHYTA: LILIOPSIDA: POALES:
OC POACEAE: HORDEUM.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-ALBOSTRIANS, CV. HAISA; TISSUE-WHITE LEAF;
RA HESS W.R., GOLZ R., BOERNER T.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: A222776; E1203984; -.
KW HYPOTHETICAL PROTEIN.
FT NON_TER 1
SQ SEQUENCE 108 AA; 12250 MW; E5FD23CE CRC32;

Query Match 56.3%; Score 54; DB 10; Length 108;
Best Local Similarity 45.5%; Pred. No. 5.16e+00;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 15 LEFTINDPP 25
QY 5 LAFKIVSOEPA 15

RESULT 7
ID 055075 PRELIMINARY; PRT; 238 AA.
AC 055075;
DT 01-JUN-1998 (TREMREL. 06, CREATED)
DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE N-CADHERIN (FRAGMENT).
OS CRICETULUS GRISEUS (CHINESE HAMSTER).
OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: MAMMALIA: EUTHERIA: RODENTIA:
OC SCOTOGNATHI: MURIDAE: CRICETINAE; CRICETULUS.
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-Ovary;
RA LEVENBERG S., SADOY E., GOICBERG P., GEIGER B.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: A003143; E1226133; -.
DE PROSITE: PS00232; CADHERIN; 2.
KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
FT NON_TER 1
SQ SEQUENCE 238 AA; 26234 MW; 0337A817 CRC32;

Query Match 56.3%; Score 54; DB 11; Length 238;
Best Local Similarity 40.0%; Pred. No. 5.16e+00;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 79 LNSKLRILSQAPS 93
QY 1 LNSKLRIVSOEPA 15

RESULT 8
ID 015855 PRELIMINARY; PRT; 878 AA.
AC 015855; 016194; 013799;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE UVOMORULIN PRECURSOR (E-CADHERIN) (ARC-1/UVOMORULIN).
GN UVOMORULIN.
OC HOMO SAPIENS (HUMAN).
OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: MAMMALIA: EUTHERIA: PRIMATES;

```

```

OC CATARRHINI; HOMINIDAE; HOMO.
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-LIVER;
RX MEDLINE: 94242050.
RA RIMM D.L., MORROW J.S.;
RT "Molecular cloning of human E-cadherin suggests a novel subdivision
of the cadherin superfamily.";
RL BIOCHEM. BIOPHYS. RES. COMMON. 200:1754-1761(1994).
RN (2)
RP SEQUENCE OF 333-472 FROM N.A.
RX MEDLINE: 94306394.
RA BECKER K.F., ATKINSON M.J., REICH U., BECKER I., NEKARDA H.,
RA STEWERT J.R., HOFER H.;
RT "E-cadherin gene mutations provide clues to diffuse type gastric
canceromas.";
RL CANCER RES. 54:3845-3852(1994).
RN (3)
RP SEQUENCE OF 261-388 FROM N.A.
RC TISSUE-LIVER;
RA FRIXEN U.H.;
RL SUBMITTED (MAR-1990) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: S72397; E136801; JOINED.
DR EMBL: L08599; G340185; -.
DR EMBL: S72492; E136801; -.
DR EMBL: S72491; E136801; JOINED.
DR PROSITE: PS00232; CADHERIN; 3.
DR PRAM: PF00028; cadherin; 5.
DR PRAM: PF01049; Cadherin_C-term; 1.
KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
FT SIGNAL 23
FT CHAIN 878
FT SIGNAL 22
SQ SEQUENCE 878 AA; 96741 MW; 8F7F0180 CRC32;

Query Match 56.3%; Score 54; DB 4; Length 878;
Best Local Similarity 53.8%; Pred. No. 5.16e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 293 NNAIVYTLSDP 305
QY 2 NSKLRIVSOEPA 14

RESULT 9
ID 089238 PRELIMINARY; PRT; 2258 AA.
AC 089238;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE LARGE PROTEIN.
OS HUMAN PARAINFLUENZA VIRUS 3.
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; PARAMYXOVIRUS.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-GP;
RX MEDLINE: 98436895.
RA OHSAWA K., YAMADA A., TAKEUCHI K., MATANABE Y., MIYATA H., SATO H.;
RT "Genetic characterization of parainfluenza virus 3 derived from
guinea pigs.";
RL J. VER. MED. SCI. 60:919-922(1998).
DR EMBL: AB012132; D1033539; -.
SQ SEQUENCE 2258 AA; 258800 MW; 8E00950F CRC32;

Query Match 56.3%; Score 54; DB 14; Length 2258;
Best Local Similarity 33.3%; Pred. No. 5.16e+00;
Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 998 LDRSVLYRIMQEPG 1012
QY 1 LNSKLRIVSOEPA 15

```

RESULT 10
ID 081081 PRELIMINARY; PRT; 2258 AA.
AC 081081;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE RNA POLYMERASE.
GN L.
OS HUMAN PARAINFLUENZA VIRUS TYPE 3.
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; PARAMYXOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-US;
RX MEDLINE: 94091054.
RA STOKES A., TIERNEY E.L., SARRIS C.M., MURPHY B.R., HALL S.L.;
RT "The complete nucleotide sequence of two cold-adapted,
RT temperature-sensitive attenuated mutant vaccine viruses (cp12 and
RT cp45) derived from the JS strain of human parainfluenza virus type 3
RT (PIV3)."
RL VIRUS RES. 30:43-52(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-US;
RA DUREIN A.D.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-US;
RA DUREIN A.D.;
RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U51116; G126234; -
DR PFAM: PF00946; Paramyx-RNA POL. 1.
SQ SEQUENCE 2258 AA; 258927 MW; 4B7A664B CRC32;

Query Match 56.38; Score 54; DB 14; Length 2258;
Best Local Similarity 33.38; Pred. No. 5.16e+00;
Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 998 LDRSVLRINNOEPG 1012
OY 1 LNSKIAFKIVSOEPA 15

RESULT 11
ID 043159 PRELIMINARY; PRT; 464 AA.
AC 043159;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DE 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE KIA0409 (FRAGMENT).
GN KIA0409.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA ISHITAMA K., NAGASE T., NAKAJIMA D., SEKI N., OHIRA M., MIYAJIMA N.,
RA TANAKA A., KOTANI H., NOMURA N., OHARA O.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AB007869; D1024586; -
FT NON_TER
SQ SEQUENCE 464 AA; 51479 MW; 4C0DD0CE CRC32;

Query Match 55.28; Score 53; DB 4; Length 464;
Best Local Similarity 54.58; Pred. No. 8.24e+00;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 416 TKLGFIVSKD 426
OY 1 LNSKIAFKIVSOEPA 15

OY 3 SKIAFKIVSOE 13
RESULT 12
ID P93205 PRELIMINARY; PRT; 775 AA.
AC P93205;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DE 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE SERINE PROTEASE, SBT2.
GN SBT2.
OS LYCOPOERSICON ESCULENTUM (TOMATO).
OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; SOLANACEAE; SOLANACEAE; SOLANACEAE; SOLANUM.
RN [1]
RP SEQUENCE FROM N.A.
RA SCHALLER A.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. VTR8;
RA WEICHTER J., AMRHEIN N., SCHALLER A.;
RT "The gene family of subtilisin-like proteases in tomato."
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: X98930; E259406; -
DR EMBL: AJ006379; E1299612; -
DR PFAM: PF00082; subtilase; 3.
DR MENDEL; 8813; LxCes; 1086; 2.
KW PROTEASE; SERINE PROTEASE.
SQ SEQUENCE 775 AA; 83115 MW; 2EBF3C5E CRC32;

Query Match 55.28; Score 53; DB 10; Length 775;
Best Local Similarity 46.78; Pred. No. 8.24e+00;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 734 LSKYVTEFKIVSROKA 748
OY 1 LNSKIAFKIVSOEPA 15

RESULT 13
ID 084185 PRELIMINARY; PRT; 1980 AA.
AC 084185;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SENDAI VIRUS (STRAIN 2) GENOME RNA 5' END (STRAIN 2).
OS HUMAN PARAINFLUENZA 1 VIRUS.
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; PARAMYXOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-2;
RX MEDLINE: 86148492.
RA SHIODA T., IMASAKI K., SHIBUTA H.;
RT "Determination of the complete nucleotide sequence of the Sendai
RT virus genome RNA and the predicted amino acid sequences of the F, HN
RT and L proteins."
RL NUCLEIC ACIDS RES. 14:1545-1563(1986).
DR EMBL: X03614; G60900; -
DR PFAM: PF00946; Paramyx-RNA POL. 1.
SQ SEQUENCE 1980 AA; 224006 MW; 4BAC22B0 CRC32;

Query Match 55.28; Score 53; DB 14; Length 1980;
Best Local Similarity 26.78; Pred. No. 8.24e+00;
Matches 4; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

Db 725 LDKQVYRVNNOEPG 739
OY 1 LNSKIAFKIVSOEPA 15

```

RESULT 14 PRELIMINARY; PRT; 2228 AA.
AC 055528;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE RN POLYMERASE PROTEIN.
GN
LN
OS SENDAI VIRUS.
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
CC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; PARAMYXOVIRUS.
RN
RP SEQUENCE FROM N.A.
RC STRAIN-OHITA.
RX MEDLINE; 98062143.
RA ITOH M., ISEGAWA Y., HOTTA H., HOMMA M.;
RT "Isolation of an avirulent mutant of Sendai virus with two amino acid
RL mutations from a highly virulent field strain through adaptation to
RL L1C-MK2 cells."
RL J. GEN. VIROL. 78:3207-3215(1997).
SO EMBL; AB005795; D1025305;
SQ SEQUENCE 2228 AA; 253058 MW; 8816E77F CRC32;

Query Match 55.2%; Score 53; DB 14; Length 2228;
Best Local Similarity 26.7%; Pred. No. 8.24e+00;
Matches 4; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

Db 973 LDKOVLRYMNOEPG 987
| : : : : : ||| :
OY 1 LNSKIAFKIVSOEPA 15

RESULT 15 PRELIMINARY; PRT; 2228 AA.
ID 055530;
AC 055530;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE RN POLYMERASE PROTEIN.
GN
LN
OS SENDAI VIRUS.
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
CC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; PARAMYXOVIRUS.
RN
RP SEQUENCE FROM N.A.
RC STRAIN-OHITA.
RX MEDLINE; 98062143.
RA ITOH M., ISEGAWA Y., HOTTA H., HOMMA M.;
RT "Isolation of an avirulent mutant of Sendai virus with two amino acid
RL mutations from a highly virulent field strain through adaptation to
RL L1C-MK2 cells."
RL J. GEN. VIROL. 78:3207-3215(1997).
SO EMBL; AB005796; D1025314;
SQ SEQUENCE 2228 AA; 253000 MW; A3058CCB CRC32;

Query Match 55.2%; Score 53; DB 14; Length 2228;
Best Local Similarity 26.7%; Pred. No. 8.24e+00;
Matches 4; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

Db 973 LDKOVLRYMNOEPG 987
| : : : : : ||| :
OY 1 LNSKIAFKIVSOEPA 15

Search completed: Fri Jun 11 17:25:14 1999
Job time : 124 secs.

```



(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_lp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 11 17:34:04 1999; Maspar time 4.88 Seconds
65.408 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-4
Description: (1-15) from US08991628.pep
Perfect Score: 108
Sequence: 1 TPMFLSRNTEGVRT 15

Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 19.257; Variance 58.572; scale 0.329

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	108	100.0	15	35	W78815	Desmoglein 3 protein	6.54e-05
2	108	100.0	15	20	W04844	Self epitope of desmo	6.54e-05
3	108	100.0	15	33	W64816	Desmoglein-3 206-220.	6.54e-05
4	108	100.0	15	19	W07908	Pemphigus vulgaris an	6.54e-05
5	108	100.0	15	99	6 R30742	Human pemphigus vulga	6.54e-05
6	88	81.5	78	21	W15489	Pemphigus foliaceus a	1.64e-02
7	61	56.5	58	2	R07999	Asparagine synthetase	1.82e+01
8	61	56.5	58	2	R07998	Asparagine synthetase	1.82e+01
9	60	55.6	16	34	W38654	S. pneumoniae aspara	2.33e+01
10	59	54.6	16	34	R86867	Rat protocadherin p05	2.98e+01
11	57	52.8	62	24	W13010	Segment of desmosomal	4.84e+01
12	57	52.8	62	24	W13009	Segment of desmosomal	4.84e+01
13	55	50.9	66	13	R69633	Human interleukin-12	7.81e+01
14	55	50.9	66	13	R69632	Human interleukin-12	7.81e+01
15	55	50.9	66	21	W12772	Human interleukin-12	7.81e+01
16	54	50.0	32	28	W33725	Mouse melanocortin-5	9.91e+01

17	54	50.0	32	28	W41067	Mouse melanocortin-5	9.91e+01
18	54	50.0	32	51	W37833	Mouse melanocortin-5	9.91e+01
19	54	50.0	32	14	R79501	Rat melanocortin rece	9.91e+01
20	54	50.0	32	22	W19705	Melanocortin-5 recept	9.91e+01
21	54	50.0	32	36	W79688	Melanocortin-5 recept	9.91e+01
22	54	50.0	43	19	W00365	Human cyclin B1.	9.91e+01
23	54	50.0	55	38	W85133	A desaturase enzyme e	9.91e+01
24	54	50.0	55	38	W84154	Human desaturase enzy	9.91e+01
25	54	50.0	60	38	W85134	A desaturase enzyme e	9.91e+01
26	54	50.0	60	38	W84155	Human desaturase enzy	9.91e+01
27	54	50.0	74	38	W84156	Human desaturase enzy	9.91e+01
28	54	50.0	74	38	W85135	A desaturase enzyme e	9.91e+01
29	53	49.1	60	4	P30206	Sequence encoded by p	1.25e+02
30	53	47.2	40	3	R10919	Human GM-CSF receptor	2.00e+02
31	51	47.2	55	2	R24017	Fusion protein GM-CSF	2.00e+02
32	51	47.2	168	34	W70991	Human class II p13 k1	2.00e+02
33	51	47.2	172	29	W38756	Phosphatidyl inositol	2.00e+02
34	50	46.3	156	2	P70062	Sequence of human gro	2.52e+02
35	50	46.3	153	1	P80877	Hook region #8 conig	2.52e+02
36	50	46.3	225	2	P70061	Sequence of human gro	2.52e+02
37	50	46.3	262	4	P40023	Fusion protein contg.	2.52e+02
38	50	46.3	405	4	P30202	Sequence encoded by p	2.52e+02
39	50	46.3	406	4	P30203	Porcine acylglycosami	2.52e+02
40	50	46.3	417	16	R79929	Farnesoid-activated r	2.52e+02
41	50	46.3	459	20	W03448	Human retinoid recept	2.52e+02
42	50	46.3	452	20	W40072	Lactobacillus amylovo	2.52e+02
43	50	46.3	720	23	W19266	Human E-cadherin prec	2.52e+02
44	50	46.3	878	15	R85487	H. pylori cytoplasmic	2.52e+02
45	50	46.3	928	22	W21017		

ALIGNMENTS

RESULT 1
ID W78815 standard; peptide: 15 AA.
AC W78815;
DT 17-NOV-1998 (first entry)
DE Desmoglein 3 protein fragment 206-220.
KW Microparticle delivery; polymeric matrix; autoantigen; tumour antigen;
KW class II associated peptide; pathogen; gene therapy; genetic disease;
KW infection; downregulation; immune response.
OS Homo sapiens.
PN WO9831398-A1.
PD 23-JUL-1998.
PF 22-JAN-1998; 001499.
PR 06-JAN-1998; US-003253.
PR 22-JAN-1997; US-787547.
PA (PANG-) PANGAEA PHARM INC.
PI Curley JM, Hedley ML, Langer RS, Lunsford LB;
DR MPI: 98-427556/36.
PT New preparations of microparticles - comprising a synthetic polymer
PT matrix and nucleic acid comprising an expression vector for use in
PT gene therapy.
PS Disclosure: Page 8; 101pp; English.
CC A microparticle preparation (MP) has been developed, consisting of
CC microparticles having a diameter of less than 100 nm. The MP comprises:
CC (a) a polymeric matrix (PM) consisting of one or more synthetic polymers
CC having a solubility in water of less than 1 mg/l; and (b) an expression
CC vector selected from RNA molecules (at least 50% of which are closed
CC circles) or circular plasmid DNA (at least 50% of which are supercoiled).
CC Also described is a MP of at most 20 microns in diameter, comprising: (a)
CC a PM; and (b) a NM comprising an expression control sequence operatively
CC linked to a coding sequence, where the coding sequence encodes an
CC expression product selected from: (1) a polypeptide at least 7 amino
CC acids in length, having a sequence identical to the sequence of: (1) a
CC fragment of a naturally-occurring mammalian protein; or (1i) a fragment
CC of a naturally-occurring protein from an infectious agent which infects
CC a mammal; (2) a peptide having a length and sequence which permits it to
CC bind to an MHC class I or II molecule; and (3) the polypeptide or the
CC peptide linked to a trafficking sequence. W69763 to W69765, and W78793
CC to W78897 are peptide fragments for use in the present invention. The
CC MPs are highly effective vehicles for the delivery of polynucleotides

CC into phagocytic cells. They can be used for gene therapy, e.g. for
 CC treating genetic diseases, infections or tumours or for downregulating
 CC an immune response.
 SQ Sequence 15 AA;

Query Match 100.0%; Score 108; DB 35; Length 15;
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 tpmflsrtngvrt 15
 |||||
 QY 1 TPMFLSRTNGVRT 15

RESULT 2
 ID W04844 standard; peptide; 15 AA.

DE Self epitope of desmoglein 3, implicated in autoimmune disease.
 KW Tolerisation; self-epitope; antigen; autoimmune disease;
 KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
 KW pemphigus vulgaris; desmoglein; multiple sclerosis;
 KW herpes simplex virus; adenovirus; phosphomannomutase;
 KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
 KW Influenza; haemagglutinin; reovirus; sigma protein.
 OS Homo sapiens.
 PN WO9627387-A1.
 PD 12-SEP-1996.
 PF 07-MAR-1996; U03182.
 PR 07-MAR-1996; US-400796.
 PA (HARD) HARVARD COLLEGE.
 PI Strominger JL, Nuchterpfennig KW;
 DR WPI; 96-425218/42.
 PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
 PT antigens - useful in disease treatment, and method for
 PT identification of other self and non-self antigens implicated in
 PT autoimmune disease.
 PS Claim 1; Page 40; 58pp; English.
 CC Pharmaceutical preparations for tolerisation to antigens comprise
 CC either an isolated human non-collagen or non-mysin basic protein
 CC (MBP) polypeptide which is capable of tolerising an individual to an
 CC autoantigen; or an isolated human pathogen polypeptide capable of
 CC tolerising an individual to that polypeptide. In both cases, the
 CC polypeptide (whether self or non-self) includes an amino acid
 CC sequence corresponding to a sequence motif for a MHC class II
 CC protein, such as HLA-DR, which is associated with a human autoimmune
 CC disease and which binds to the polypeptide to activate autoreactive
 CC T-cells in individuals with the autoimmune disease. This peptide is
 CC derived from the human desmoglein 3 protein (amino acids 206-220)
 CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
 CC derived from the human desmoglein protein are described in W04841-47.
 SQ Sequence 15 AA;

Query Match 100.0%; Score 108; DB 20; Length 15;
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 tpmflsrtngvrt 15
 |||||
 QY 1 TPMFLSRTNGVRT 15

RESULT 3
 ID W64816 standard; peptide; 15 AA.

DE Desmoglein-3 206-220.
 KW autoantigen; autoimmune disease; MHC.
 OS Homo sapiens.
 PN US5783567-A.
 PD 21-JUL-1998.
 PF 22-JAN-1997; 787547.

PR 22-JAN-1997; US-787547.
 PA (PANG-) PANGAEA PHARM INC.
 PI Curley JM, Hedley ML, Langer RS;
 DR WPI; 98-427077/36.

PT Microparticle encapsulated nucleic acids - for recombinant
 PT expression of proteins e.g. in gene therapy
 PS Disclosure; Column 4; 42pp; English.
 CC The patent describes a new preparation of microparticles each
 CC comprising a polymeric matrix and a nucleic acid. The polymeric
 CC matrix consists of one or more synthetic polymers having a solubility
 CC in water of less than 1 mg/l (e.g. poly-lactic-co-glycolic acid);
 CC and at least 90% of the microparticles have a diameter of less than
 CC 100 microns. The microparticles are useful for the delivery of nucleic
 CC acids to phagocytic cells. In one embodiment the microparticles are
 CC less than 20 microns in diameter and the nucleic acid (preferably in
 CC closed circular form) includes an expression control sequence
 CC operatively linked to a coding sequence, where the expression product
 CC of the coding sequence is a polypeptide having a length and a sequence
 CC which permits it to bind to an MHC class I or II molecule. The
 CC expression product is thus an effective stimulator of an immune
 CC response in mammals. The present sequence, an antigenic portion of
 CC desmoglein 3, is an example of an MHC class II peptide which can be
 CC expressed by the nucleic acid. It is associated with pemphigus
 CC vulgaris.
 SQ Sequence 15 AA;

Query Match 100.0%; Score 108; DB 33; Length 15;
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 tpmflsrtngvrt 15
 |||||
 QY 1 TPMFLSRTNGVRT 15

RESULT 4
 ID W07908 standard; protein; 614 AA.

DE 29-JAN-1997 (first entry)
 DE Pemphigus vulgaris antigen protein extracellular region.
 KW Autoantibody; immunoglobulin G; IgG1; fusion protein; diagnosis;
 KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;
 KW dermatology.
 OS Homo sapiens.
 PN J08188540-A.
 PD 23-JUL-1996.
 PF 30-JUN-1995; 165632.
 PR 30-JUN-1994; JP-173291.
 PA (NISH/) NISHIKAWA T.
 DR WPI; 96-388562/39.
 PT Fused protein recognised by pemphigus vulgaris auto-antibody -
 PT useful to treat and diagnose pemphigus vulgaris
 PS Claim 1; Page 7-9; 9pp; Japanese.
 CC W07908 represents the human pemphigus vulgaris (PV) antigen
 CC extracellular region. The PV antigen is produced in patients with
 CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare
 CC relapsing disease causing suprabasal, intra-epidermal bullae
 CC (vesicles) of the skin and mucous membranes, which is fatal if
 CC untreated. The PV antigen was fused to a human IgG1 hinge region
 CC and the resulting fusion protein is useful to treat or diagnose
 CC pemphigus vulgaris.
 SQ Sequence 614 AA;

Query Match 100.0%; Score 108; DB 19; Length 614;
 Best Local Similarity 100.0%; Pred. No. 6.54e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 205 tpmflsrtngvrt 219
 |||||
 QY 1 TPMFLSRTNGVRT 15

RESULT 5

FT binding_site 1..4
 PN /label- putative glutamine binding site
 MO9013633-A.
 PD 15-NOV-1990.
 PF 02-MAY-1990: U02443.
 PR 03-MAY-1989: US-347302.
 PR 26-APR-1990: US-514816.
 PA (UYRO-) ROCKFELLER UNIV.
 PI Coruzzi GM, Tsai FY;
 DR N-PSDB: 006598.
 PT plant asparagine synthetase - includes DNA expression system and
 PT transgenic plants
 PS Disclosure; Fig 2A; 91pp; English.
 CC The DNA sequence encoding this protein was isolated from cDNA
 CC clones selected from a pea node cDNA library from the "Sparkle"
 CC variety of *P. sativum*. Human AS cDNA was used as a probe.
 CC The protein is produced by expression vectors containing the ASI
 CC coding sequence. Recombinant AS can be used to engineer herbicide
 CC resistance, as a dominant selectable marker, to select novel
 CC herbicides or compounds useful for synchronising plant cells in
 CC culture, etc.
 CC See also Q06599, Q06622 and Q06623.
 SQ Sequence 586 AA;

Query Match 56.5%; Score 61; DB 2; Length 586;
 Best Local Similarity 77.8%; Pred. No. 1.82e+01;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 323 tpmflmrk 331
 1 tpmflmrk 9

QY 1 tpmflmrk 9

RESULT 9
 ID W8654 standard; Protein; 163 AA.
 AC W8654;
 DT 09-NOV-1998 (first entry)
 DE *S. pneumoniae* aspartate ammonia ligase.
 KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;
 KW immunological response; inoculation; antibody production; inhibitor;
 KW T cell immune response; antimicrobial compound; bacterial adhesion;
 KW extracellular matrix protein; protein-mediated cell invasion; wound;
 KW pathogenesis.
 OS Streptococcus pneumoniae.
 PN MO9743303-A1.
 PD 20-NOV-1997.
 PF 14-MAY-1997: U07950.
 PR 14-MAY-1996: US-017670.
 PA (SMIR) SMITHKLINE BEECHAM CORP.
 PA (SMIR) SMITHKLINE BEECHAM PLC.
 PI Black Mt. Hodgson JE, Knowles DDC, Nicholas RO,
 PI Stodola RK;
 DR WPI: 98-008793/01.
 DR N-PSDB: T98699.
 PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
 PT diagnosing anti-microbial agents for treatment of bacterial
 PT infections
 PS Claim 12: Page 407: 483pp; English.
 CC This sequence represents a Streptococcus pneumoniae protein that, based
 CC on homology with an Escherichia coli protein, is an aspartate ammonia
 CC ligase, and is encoded by a DNA sequence of the invention.
 CC The DNA sequences were isolated from Streptococcus pneumoniae strain
 CC 1100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
 CC invention can be used to identify compounds which interact with and
 CC inhibit or activate the activity of the proteins. Antagonists can be
 CC used to treat diseases caused by *S. pneumoniae* proteins, through genetic
 CC immunisation. They can also be used to induce an immunological response
 CC in a mammal by inoculation with the *S. pneumoniae* proteins or delivery
 CC of the encoding nucleic acids in a vector adequate to produce antibody
 CC and/or T cell immune responses to protect the animal from disease. The
 CC proteins can also be used to identify antimicrobial compounds which are
 CC capable of inhibiting their bioactivity. In particular the proteins of

CC the invention can be used to prevent adhesion of bacteria to mammalian
 CC extracellular matrix proteins on in-dwelling devices or in wounds, to
 CC block protein-mediated mammalian cell invasion, and to block the normal
 CC progression of pathogenesis in infections initiated other than by the
 CC implantation of in-dwelling devices or other surgical techniques.
 SQ Sequence 163 AA;

Query Match 55.6%; Score 60; DB 34; Length 163;
 Best Local Similarity 60.0%; Pred. No. 2.33e+01;
 Matches 9; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

Db 115 amflmrk 129
 2 tpmflmrk 15

QY 2 tpmflmrk 15

RESULT 10
 ID R8687 standard; Protein; 616 AA.
 AC R8687;
 DT 27-AUG-1996 (first entry)
 DE Rat protocadherin pc5.
 KW Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion;
 KW catenin; therapy.
 OS Rattus rattus.
 PN MO9600289-A1.
 PD 04-JAN-1996.
 PF 26-JUN-1995: U08071.
 PR 27-JUN-1994: US-268161.
 PA (DOHE-) DOHERTY EYE INST.
 PI Suzuki S;
 DR WPI: 96-068873/07.
 DR N-PSDB: T03574.
 PT Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
 PT pc5 involved in cell-cell adhesion and regulation activities
 PS Claim 17: Page 130-132: 146pp; English.
 CC R8685-R8687 represent the sequences for three protocadherins. This
 CC sequence represents the rat protocadherin pc5. These sequences are
 CC related to cadherin, and possess cell adhesive ability. Cadherins are
 CC glycosylated integral membrane proteins that are involved in cell-cell
 CC adhesion. Cadherins are composed of an N-terminal extracellular domain
 CC which consists of 5 unique subdomains, a membrane spanning domain, and a
 CC C-terminal cytoplasmic domain. The cytoplasmic domain interacts with the
 CC cytoskeleton through catenins and other cytoskeleton associated proteins.
 CC The cytoplasmic domain is not present in all cadherins, but in those
 CC which possess it, it is essential for the cadherins adhesive function.
 CC The cadherins which do not possess a cytoplasmic domain appear to
 CC function via a different method from those with a cytoplasmic domain.
 CC These sequences were isolated using primers 1 and 2 (see T03575 and
 CC T03576). The proteins may have regulatory functions in the cell, as well
 CC as the cell-cell adhesive properties. Antibodies produced against these
 CC sequences are useful for modulating the binding activity of these
 CC protocadherins, and can be used therapeutically.
 SQ Sequence 616 AA;

Query Match 54.6%; Score 59; DB 17; Length 616;
 Best Local Similarity 61.5%; Pred. No. 2.98e+01;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 85 tpmflmrk 97
 1 tpmflmrk 13

QY 1 tpmflmrk 13

RESULT 11
 ID W13010 standard; Protein; 263 AA.
 AC W13010;
 DT 21-NOV-1997 (first entry)
 DE Segment of desmosomal cadherin, desmoglein Dsg2; cell; surface; epithelial;
 KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;
 KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;
 KW micrometastasis; separation; enrichment; targeted delivery;
 KW metastatic.
 OS Homo sapiens.

ID	RESULT	ID	RESULT	ID	RESULT
D6	160 pfvflnkdgtel 171	AC	R69633; standard; Protein; 660 AA.	AC	R69633; standard; Protein; 660 AA.
QY	2 PMFLSRNTGEV 13	FT	05-SEP-1995 (first entry)	FT	05-SEP-1995 (first entry)
		FT	Human interleukin-12 receptor	FT	Human interleukin-12 receptor
		FT	Interleukin-12 receptor; IL-12; immune suppression;	FT	Interleukin-12 receptor; IL-12; immune suppression;
		FT	Immunosuppressive; graft-versus-host reaction; allograft rejection	FT	Immunosuppressive; graft-versus-host reaction; allograft rejection
		FT	Inflammation; autoimmune disease.	FT	Inflammation; autoimmune disease.
		FT	Homo sapiens.	FT	Homo sapiens.
		FT	Key	FT	Key
		FT	peptide	FT	peptide
		FT	Location/Qualifiers	FT	Location/Qualifiers
		FT	1..23	FT	1..23
		FT	/label= Sig_peptide	FT	/label= Sig_peptide
		FT	24..540	FT	24..540
		FT	/note= "extracellular region"	FT	/note= "extracellular region"
		FT	52..64	FT	52..64
		FT	/note= "sequence motif of cytokine receptor superfamily Cys52..Cys62Sw"	FT	/note= "sequence motif of cytokine receptor superfamily Cys52..Cys62Sw"
		FT	121..123	FT	121..123
		FT	/label= N-linked_glycosylation	FT	/label= N-linked_glycosylation
		FT	222..226	FT	222..226
		FT	/note= "cytokine receptor superfamily motif (W22SKWS)"	FT	/note= "cytokine receptor superfamily motif (W22SKWS)"
		FT	329..331	FT	329..331
		FT	/label= N-linked_glycosylation	FT	/label= N-linked_glycosylation
		FT	346..348	FT	346..348
		FT	/label= N-linked_glycosylation	FT	/label= N-linked_glycosylation
		FT	352..354	FT	352..354
		FT	/label= N-linked_glycosylation	FT	/label= N-linked_glycosylation
		FT	442..444	FT	442..444
		FT	/label= N-linked_glycosylation	FT	/label= N-linked_glycosylation
		FT	456..458	FT	456..458
		FT	/label= N-linked_glycosylation	FT	/label= N-linked_glycosylation
		FT	541..570	FT	541..570
		FT	/label= Transmembrane_region	FT	/label= Transmembrane_region
		FT	571..662	FT	571..662
		FT	/label= Cytoplasmic_tail_region	FT	/label= Cytoplasmic_tail_region
		FT	577..584	FT	577..584
		FT	/label= Cytoplasmic_tail	FT	/label= Cytoplasmic_tail
		FT	/note= "conserved area of cytoplasmic tail"	FT	/note= "conserved area of cytoplasmic tail"
		FT	618..629	FT	618..629
		FT	/label= Cytoplasmic_tail	FT	/label= Cytoplasmic_tail
		FT	/note= "conserved area of cytoplasmic tail"	FT	/note= "conserved area of cytoplasmic tail"
		FT	region	FT	region
		FT	EP-638644-A.	FT	EP-638644-A.
		FT	15-FEB-1995.	FT	15-FEB-1995.
		FT	08-JUL-1994; 110657.	FT	08-JUL-1994; 110657.
		FT	19-JUL-1993; US-094649.	FT	19-JUL-1993; US-094649.
		FT	19-JUL-1993; US-094713.	FT	19-JUL-1993; US-094713.
		FT	31-MAY-1994; US-248532.	FT	31-MAY-1994; US-248532.
		FT	(HOFF) HOFFMANN LA ROCHE & CO AG F.	FT	(HOFF) HOFFMANN LA ROCHE & CO AG F.
		FT	Chazzone Ra, China AO, Gubler UA, Trullit TP;	FT	Chazzone Ra, China AO, Gubler UA, Trullit TP;
		FT	WP1; 95-076349/11.	FT	WP1; 95-076349/11.
		FT	DNA encoding a low affinity interleukin-12 receptor - used to	FT	DNA encoding a low affinity interleukin-12 receptor - used to
		FT	bind or scavenge IL-12 to cause immune suppression, e.g. to	FT	bind or scavenge IL-12 to cause immune suppression, e.g. to
		FT	suppress graft-vs-host reaction, allograft rejection or	FT	suppress graft-vs-host reaction, allograft rejection or
		FT	inflammation, and to treat autoimmune conditions	FT	inflammation, and to treat autoimmune conditions
		FT	Claim 4; Page 31-33; 61pp; English.	FT	Claim 4; Page 31-33; 61pp; English.
		FT	A cDNA library of PHA-activated peripheral blood mononuclear cells	FT	A cDNA library of PHA-activated peripheral blood mononuclear cells
		FT	in vector pEF-BOS was screened for interleukin-12 (IL-12) receptor	FT	in vector pEF-BOS was screened for interleukin-12 (IL-12) receptor
		FT	cDNAs by panning. One isolated cDNA (083844) encoded a 662-amino	FT	cDNAs by panning. One isolated cDNA (083844) encoded a 662-amino
		FT	acid low affinity IL-12 receptor (R69632). Another cDNA contained	FT	acid low affinity IL-12 receptor (R69632). Another cDNA contained
		FT	an additional 202 bp of 3'UTR; the deduced 660-amino acid sequence	FT	an additional 202 bp of 3'UTR; the deduced 660-amino acid sequence
		FT	(R69633) showed an altered C-terminal sequence compared to R69632	FT	(R69633) showed an altered C-terminal sequence compared to R69632
		FT	(Recombinant IL-12 receptor was expressed in COS cells, and can be	FT	(Recombinant IL-12 receptor was expressed in COS cells, and can be
		FT	used for therapeutic or diagnostic purposes.	FT	used for therapeutic or diagnostic purposes.
		FT	Sequence 660 AA.	FT	Sequence 660 AA.
		FT	Query Match	FT	Query Match
		FT	50.9%; Score 55; DB 13; Length 660;	FT	50.9%; Score 55; DB 13; Length 660;
		FT	Best Local Similarity 53.8%; Pred. NO. 7.81e+01;	FT	Best Local Similarity 53.8%; Pred. NO. 7.81e+01;

Matches 7: Conservative 2: Mismatches 4: Indels 0: Gaps 0:
 Db 14 1fl1srqgaact 26
 :|||||: ||
 QY 3 MFLSRNTGEVRT 15

RESULT 14
 ID R69632 standard; Protein: 662 AA.
 AC R69632;
 DT 05-SEP-1995 (first entry)

DE Human interleukin-12 receptor.
 DE Interleukin-12 receptor; IL-12; immune suppression;
 KW immunosuppressive; graft-versus-host reaction; allograft rejection;
 KW inflammation; autoimmune disease.
 OS Homo sapiens.

FT Key Location/Qualifiers
 FT peptide 1..23
 FT /label- Sig-peptide
 FT region 24..540
 FT /note- "extracellular region"
 FT region 52..64
 FT /note- "sequence motif of cytokine receptor
 FT superfamily Cys52..Cys62Sw"
 FT modified_site 121..123
 FT /label- N-linked-glycosylation
 FT modified_site 222..226
 FT /note- "cytokine receptor superfamily motif
 FT (W22SKWS)"
 FT modified_site 329..331
 FT /label- N-linked-glycosylation
 FT modified_site 346..348
 FT /label- N-linked-glycosylation
 FT modified_site 352..354
 FT /label- N-linked-glycosylation
 FT modified_site 442..444
 FT /label- N-linked-glycosylation
 FT modified_site 456..458
 FT /label- N-linked-glycosylation
 FT modified_site 541..570
 FT /label- Transmembrane_region
 FT region 571..662
 FT /label- Cytoplasmic-tail_region
 FT region 577..584
 FT /label- Cytoplasmic-tail
 FT /note- "conserved area of cytoplasmic tail"
 FT region 618..629
 FT /label- Cytoplasmic-tail
 FT /note- "conserved area of cytoplasmic tail"

FN EP-638644-A.
 PD 15-FEB-1995.
 PF 08-JUL-1994; 110657.
 PR 19-JUL-1993; US-094649.
 PR 19-JUL-1993; US-094713.
 PR 31-MAY-1994; US-248532.
 PI (HOF) HOFFMANN LA ROCHE & CO AG F.
 PI Chizomite RA, Chua AO, Gubler UA, Trullitt TP;
 DR N-PSDB; 083844.
 DT WPI: 95-076349/11.
 DR N-PSDB; 083844.
 PT bind or scavenge IL-12 to cause immune suppression, e.g. to
 PT suppress graft-vs-host reaction, allograft rejection or
 PT inflammation, and to treat autoimmune conditions
 PS Claim 4; Page 24-27; 61pp; English.
 CC A cDNA library of PHA-activated peripheral blood mononuclear cells
 CC in vector pEF-BOS was screened for interleukin-12 (IL-12) receptor
 CC cDNAs by panning. An isolated cDNA was sequenced (083844); it
 CC encoded a 662-amino acid low affinity IL-12 receptor (R69632).
 CC Recombinant IL-12 receptor was expressed in COS cells, and can be
 CC used for therapeutic or diagnostic purposes.
 SQ Sequence 662 AA;

Query Match

50.9%; Score 55; DB 13; Length 662;

Best Local Similarity 53.8%; Pred. No. 7.81e+01;
 Matches 7: Conservative 2: Mismatches 4: Indels 0: Gaps 0:

Db 14 1fl1srqgaact 26
 :|||||: ||
 QY 3 MFLSRNTGEVRT 15

RESULT 15
 ID W12772 standard; Protein: 662 AA.
 AC W12772;
 DT 12-MAY-1997 (first entry)

DE Human interleukin-12 beta-1 receptor.
 DE Interleukin-12 beta-1 receptor; IL-12; autoimmune disease;
 KW rheumatoid arthritis; inflammatory bowel disease;
 KW multiple sclerosis.
 OS Homo sapiens.

FT Key Location/Qualifiers
 FT peptide 1..20
 FT /label- Sig-peptide
 FT /note- "signal peptide cleavage site alternatively
 FT follows Ala-23 if Cys-24"
 FT domain 24..540
 FT /label- Extracellular_domain
 FT domain 540..570
 FT /label- Transmembrane_domain
 FT domain 571..662
 FT /label- Cytoplasmic-tail
 FT region 52..64
 FT /label- Cytokine-receptor_motif
 FT /note- "cytokine receptor superfamily motif
 FT (Cys52..Cys62Sw)"
 FT region 222..226
 FT /label- Cytokine-receptor_motif
 FT /note- "cytokine receptor superfamily motif
 FT (W22SKWS)"
 FT modified_site 121..123
 FT /label- N-linked-glycosylation_site
 FT modified_site 329..331
 FT /label- N-linked-glycosylation_site
 FT modified_site 346..348
 FT /label- N-linked-glycosylation_site
 FT modified_site 352..354
 FT /label- N-linked-glycosylation_site
 FT modified_site 442..444
 FT /label- N-linked-glycosylation_site
 FT modified_site 456..458
 FT /label- N-linked-glycosylation_site
 FT /label- N-linked-glycosylation_site

FN EP-759466-A2.
 PD 26-FEB-1997.
 PF 23-JUL-1996; 111807.
 PR 01-AUG-1995; US-001701.
 PR 30-MAY-1996; US-018674.
 PI (HOF) HOFFMANN LA ROCHE & CO AG F.
 PI Gubler UA, Presky DH;
 DR N-PSDB; T59732.
 DT WPI: 97-147515/14.
 DR N-PSDB; T59732.
 PT New interleukin-12 beta-2 receptor and high binding affinity
 PT complexes - have a high affinity for interleukin-12, and are used to
 PT treat autoimmune diseases
 PS Claim 8; Page 37-41; 53pp; English.
 CC Human interleukin-12 (IL-12) receptor beta-1 receptor (W12772)
 CC has a low binding affinity for IL-12, but when complexed with an
 CC IL-12 beta-2 receptor (see also W12771), forms a complex with a
 CC high binding affinity for IL-12. Its amino acid sequence was
 CC deduced from a cDNA clone (T59732) obtd. from human lymphoblasts.
 CC IL-12 receptor beta-1 can be expressed on the surface of
 CC transformed host cells as a complex with co-expressed IL-12 receptor
 CC beta-2, and used in therapeutic compsns., pref. with at least 1
 CC cytokine antagonist, to treat autoimmune dysfunctions such as
 CC rheumatoid arthritis, inflammatory bowel disease and multiple
 CC sclerosis. The receptor protein or complex can also be used to
 CC detect (ant)agonists of IL-12 activity.

5Q Sequence 662 AA;

Query Match 50.9%; Score 55; DB 21; Length 662;
 Best Local Similarity 53.8%; Pred. No. 7.8le+01;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 14 1fllsrggact 26

Oy 3 MFLSRNTGEVPT 15

Search completed: Fri Jun 11 17:35:53 1999
 Job time : 109 secs.

Db 206 TPFMLSRNTGEVRT 220
 |||||
 Oy 1 TPFMLSRNTGEVRT 15

RESULT 2

ENTRY IJHUG1 #type complete
 TITLE desmoglein 1 precursor - human
 ALTERNATE_NAMES desmosomal glycoprotein I
 ORGANISM #formal_name Homo sapiens #common_name man
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 26-Feb-1998

ACCESSIONS S16906; A39706; A61254; A61279; S16158

REFERENCE S16906
 #authors Buxton, R.S.
 #submission submitted to the EMBL Data Library, November 1990
 #accession S16906
 #molecule_type mRNA

REFERENCE A39706
 #cross-references EMBL:X56654; NID:g30505; PID:g30506
 #authors Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Ataliotis, P.; Poynter, D.; Arnemann, J.; Rutman, A.J.; Pidsley, S.C.; Matt, F.M.; Rees, D.A.; Buxton, R.S.; Magee, A.I.
 #journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:4796-4800
 #title Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions, is related to the cadherin family of cell adhesion molecules.

#cross-references MUID:91271279
 #accession A39706
 #molecule_type mRNA

REFERENCE A61254
 #cross-references GB:X56654
 #authors Nilles, L.A.; Parry, D.A.D.; Powers, E.E.; Angst, B.D.; Wagner, R.M.; Green, K.J.
 #journal J. Cell Sci. (1991) 99:809-821
 #title Structural analysis and expression of human desmoglein: a cadherin-like component of the desmosome.

#accession A61254
 #molecule_type mRNA

REFERENCE A61279
 #cross-references MUID:92175187
 #authors Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arnemann, J.; Rees, D.A.; King, I.A.; Magee, A.I.
 #journal Biochem. Soc. Trans. (1991) 19:1060-1064
 #title Desmosomal glycoproteins I, II and III: novel members of the cadherin superfamily.

#cross-references MUID:92175187
 #accession A61279
 #status not compared with conceptual translation
 #molecule_type mRNA

GENETICS
 #residues 1-55 ##label WH3

FEATURE
 1-23 #domain signal sequence #status predicted #label SIG
 24-49 #domain propeptide #status predicted #label PRO
 50-1049 #product desmoglein #status predicted #label MAT
 50-548 #domain extracellular #status predicted #label EXT
 52-157 #domain cadherin repeat homology #label CR1
 160-269 #domain cadherin repeat homology #label CR2
 272-385 #domain cadherin repeat homology #label CR3
 392-493 #domain cadherin repeat homology #label CR4
 509-530 #region serine/threonine-rich
 549-569 #domain transmembrane #status predicted #label TM
 572-1049 #domain intracellular #status predicted #label INT

840-869 #domain desmoglein repeat #label DG1
 870-899 #domain desmoglein repeat #label DG2
 900-927 #domain desmoglein repeat #label DG3
 928-956 #domain desmoglein repeat #label DG4
 969-1019 #region glycine/serine-rich
 110,180 #binding_site carboxydrate (Asn) (covalent) #status predicted

SUMMARY #length 1049 #molecular-weight 113715 #checksum 4482

Query Match 81.5%; Score 88; DB 1; Length 1049;
 Best Local Similarity 66.7%; Pred. No. 1.87e-06;
 Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 206 SPMFIRNRTGEVRT 220
 |||||
 Oy 1 TPFMLSRNTGEVRT 15

RESULT 3

ENTRY IJHUG1 #type complete
 TITLE desmoglein 1 precursor - bovine
 ALTERNATE_NAMES desmoglein BSGM
 ORGANISM #formal_name Bos primigenius taurus #common_name cattle
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 05-Sep-1997

ACCESSIONS S14603; A38872; A37785; S38721; A48173; S24412

REFERENCE S14603
 #authors Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
 #submission submitted to the EMBL Data Library, March 1991
 #description Complete sequence of the desmoglein precursor and evidence for the existence of different desmoglein genes expressed in cell type-specific patterns.

#accession S14603
 #molecule_type mRNA

REFERENCE A38872
 #cross-references EMBL:X58466; NID:g306; PID:g307
 #authors Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.
 #journal Eur. J. Cell Biol. (1991) 55:200-208
 #title Complete amino acid sequence of the epidermal desmoglein precursor polypeptide and identification of a second type of desmoglein gene.

#cross-references MUID:92037656
 #accession A38872
 #molecule_type mRNA

REFERENCE A37785
 #cross-references GB:S64268; GB:S64270
 #authors Goodwin, L.; Hill, J.E.; Raynor, K.; Rasz, L.; Manabe, M.; Cowlin, P.
 #journal Biochem. Biophys. Res. Commun. (1990) 173:1224-1230
 #title Desmoglein shows extensive homology to the cadherin family of cell adhesion molecules.

#cross-references MUID:91097553
 #accession A37785
 #molecule_type mRNA

REFERENCE S38721
 #cross-references GB:M58165; NID:9162966; PID:9552318
 #authors Zimbelmann, R.
 #submission submitted to the EMBL Data Library, February 1991
 #accession S38721
 #molecule_type mRNA

REFERENCE A48173
 #cross-references EMBL:X57784; NID:g436061; PID:g436062
 #authors Koch, P.J.; Walsh, M.J.; Schmelz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
 #journal Eur. J. Cell Biol. (1990) 53:1-12
 #title Identification of desmoglein, a constitutive desmosomal glycoprotein, as a member of the cadherin family of cell adhesion molecules.

#cross-references MUID:91168965
#accession A48173
#molecule_type mRNA
#residues 44-1001, 'AOPPSAT' #label KO3
#cross-references GB:X57784
#note this sequence has been revised in references A38872 and S38721

GENETICS
#gene DSG1
#classification #superfamily cadherin; cadherin repeat homology calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein
#keywords

FEATURE
1-23
24-49 #domain signal sequence #status predicted #label SRC\
50-1043 #domain propeptide #status predicted #label PRO\
50-548 #product desmoglein #status predicted #label MAT\
52-157 #domain extracellular #status predicted #label EXT\
160-269 #domain cadherin repeat homology #label CR1\
332-481 #domain cadherin repeat homology #label CR2\
349-574 #domain cadherin repeat homology #label CR4\
575-1043 #domain transmembrane #status predicted #label TM\
846-875 #domain intracellular #status predicted #label INT\
876-905 #domain desmoglein repeat #label DG1\
906-933 #domain desmoglein repeat #label DG2\
934-962 #domain desmoglein repeat #label DG3\
963-1012 #domain desmoglein repeat #label DG4\
110 #region glycyne/serine-rich
#binding_site carbohydrate (Asn) (covalent) #status experimentally
#binding_site carbohydrate (Asn) (covalent) #status predicted
180,496 #predicted

SUMMARY
#length 1043 #molecular-weight 112242 #checksum 6897

Query Match
Best Local Similarity 75.0%; Score 81; DB 1; Length 1043;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 206 SPMFIRRYTGERT 220
OY 1 TPWFLLSRNGEVRT 15
:|||||:|||||:
:|||||:|||||:

RESULT 4
ENTRY S55982 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) ASN2
#residues 44-1001, 'AOPPSAT' #label KO3
#cross-references GB:X57784
#note this sequence has been revised in references A38872 and S38721

ALTERNATE_NAMES
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 17-Jul-1998
#accession S55982; S64433
#cross-references EMBL:X83099; NID:g642340; PID:g642347

ACCESSIONS
#authors van Dyck, L.; Goffeau, A.
#submission submitted to the EMBL Data Library, December 1994
#description Genes for an asn synthase, a GUG-motif nucleoporin and a putative homeobox-domain protein are identified on a 18.3 kb segment of the yeast chromosome VII also carrying MPT1, PPT1, tree new ORFs, remnants of Ty and three tRNA genes.

FEATURE
#accession S55982
#molecule_type DNA
#residues 1-572 #label VAN
#cross-references EMBL:X83099; NID:g642340; PID:g642347

REFERENCE
#authors Van Dyck, L.; Skala, J.; de Wergifosse, P.; Purnelle, B.; Talla, E.; Nawrocki, A.; Del Bino, S.; Goffeau, A.
#submission submitted to the Protein Sequence Database, May 1996
#accession S64433
#molecule_type DNA
#residues 1-572 #label VAN
#cross-references EMBL:272909; NID:g1323202; PID:e243463; PID:g1323203; MIPS:YGR124w
#experimental_source strain S288C

GENETICS
#gene SGD:ASN2
#cross-references SGD:S0003356; MIPS:YGR124w
#map_position 7R
#classification #superfamily asparagine synthase (glutamine-hydrolyzing) asparagine biosynthesis; ligase
#keywords

FEATURE
2
#active_site Cys (amide transfer) #status predicted

SUMMARY
#length 572 #molecular-weight 64593 #checksum 9651

Query Match
Best Local Similarity 60.2%; Score 65; DB 2; Length 572;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 346 TPWFLLSRK 354
OY 1 TPWFLLSRN 9
:|||||:
:|||||:

RESULT 5
ENTRY S52694 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) ASN1
#residues 1-572 #label DAN
#cross-references EMBL:248675; NID:g1163061; PID:g747902

ALTERNATE_NAMES
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 17-Jul-1998
#accession S52694; S69033
#cross-references EMBL:248675; NID:g1163061; PID:g747902

ACCESSIONS
#authors Darg, V.D.; Bojoltin-Fukuhara, M.; Daignan-Fornier, B.
#submission submitted to the EMBL Data Library, March 1995
#description Multiple regulatory systems control expression of the Saccharomyces cerevisiae ASN1 and ASN2 genes at the transcription level.

FEATURE
#accession S52694
#molecule_type DNA
#residues 1-572 #label DAN
#cross-references EMBL:248675; NID:g1163061; PID:g747902

REFERENCE
#authors Fulton, L.
#submission submitted to the EMBL Data Library, November 1995
#description The sequence of S. cerevisiae cosmid 9659.
#accession S69033
#molecule_type DNA
#residues 1-572 #label FUT
#cross-references EMBL:U04829; NID:g1066476; PID:g1066479; MIPS:YPR145w

GENETICS
#gene SGD:ASN1
#cross-references SGD:S0006349; MIPS:YPR145w
#map_position 16R
#classification #superfamily asparagine synthase (glutamine-hydrolyzing) asparagine biosynthesis; ligase
#keywords

FEATURE
2
#active_site Cys (amide transfer) #status predicted

SUMMARY
#length 572 #molecular-weight 64470 #checksum 8736

Query Match
Best Local Similarity 60.2%; Score 65; DB 2; Length 572;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 347 TPWFLLSRK 355
OY 1 TPWFLLSRN 9
:|||||:
:|||||:

RESULT 6
ENTRY S32892 #type complete
TITLE hypothetical protein 6 precursor - Salmonella typhimurium
#residues 1-572 #label VAM
#cross-references EMBL:272909; NID:g1323202; PID:e243463; PID:g1323203; MIPS:YGR124w
#experimental_source strain S288C

#authors Friedrich, M.J.; Kinsey, N.E.; Vlla, J.; Kadner, R.J.
#journal Mol. Microbiol. (1993) 8:543-558
#title Nucleotide sequence of a 13.9kb segment of the 90kb virulence plasmid of *Salmonella typhimurium*: the presence of flmbrial biosynthetic genes.
#cross-references M01D:93316852
#accession S32892
#status preliminary
#molecule-type DNA
#residues 1-295 ##label FRI
#cross-references EMBL:L08613
SUMMARY #length 295 #molecular-weight 31162 #checksum 8332

Query Match 58.3%; Score 63; DB 2; Length 295;
Best Local Similarity 80.0%; Pred. No. 4,47e-01;
Matches 12; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

DB 132 TLMFSLR-TGEVRT 145
1 ||||| |||||
QY 1 TPFLLSRNTGEVRT 15

RESULT 7
ENTRY S56364 #type complete
TITLE inner membrane copper tolerance protein cycZ - *Escherichia coli*

ALTERNATE_NAMES thiol:disulfide interchange protein dsbd
ORGANISM #formal_name *Escherichia coli*
DATE 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Nov-1997
S56364; 141028; 141037; S42064; F65223; S57220; S47295
S56314
#cross-references M01D:95354659
#accession S57220
#status preliminary
#molecule-type DNA
#residues 1-565 ##label BUR
#cross-references EMBL:U014003; NID:91263172; PID:9535980
#note the nucleotide sequence was submitted to the EMBL Data Library, August 1994

REFERENCE
#authors Fong, S.T.; Camakaris, J.; Lee, B.T.
#journal Mol. Microbiol. (1995) 15:1127-1137
#title Molecular genetics of a chromosomal locus involved in copper tolerance in *Escherichia coli* K-12.
#cross-references M01D:95349397
#accession 141028
#status translated from GB/EMBL/DBJ
#molecule-type DNA
#residues 'M', 78-565 ##label RMS
#cross-references EMBL:X77707; NID:9535290; PID:9581055
#note in this report, the codon GNG for Val-77 was interpreted as a start codon

REFERENCE
#authors Crooke, H.; Cole, J.
#journal Mol. Microbiol. (1995) 15:1139-1150
#title The biogenesis of C-type cytochromes in *Escherichia coli* requires a membrane-bound protein, DlpZ, with a protein disulphide isomerase-like domain.
#cross-references M01D:95349398
#accession 141037
#status translated from GB/EMBL/DBJ
#molecule-type DNA
#residues 'M', 78-565 ##label RS2
#cross-references EMBL:X77707; NID:9871027; PID:9871029
#note in this report, the codon GNG for Val-77 was interpreted as a start codon

REFERENCE S42063

#authors Crooke, H.R.; Cole, J.A.
#submission submitted to the EMBL Data Library, February 1994
#description The biogenesis of C-type cytochromes in *Escherichia coli* requires an integral membrane protein with a protein disulphide isomerase like domain.
#accession S42064
#molecule-type DNA
#residues 1-328,'V',330-565 ##label CRO
#cross-references EMBL:X77707
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of *Escherichia coli* K-12.
#cross-references M01D:97426517
#accession F65223
#status nucleic acid sequence not shown; translation not shown
#molecule-type DNA
#residues 1-565 ##label BLAT
#cross-references GB:AE000486; GB:U00096; NID:91790574; PID:91790578; UWCP:D4136

REFERENCE
#experimental_source strain K-12, substrain MG1655
S57220
#authors Missiakas, D.; Schwager, F.; Raina, S.
#journal EMBO J. (1995) 14:3415-3424
#title Identification and characterization of a new disulfide isomerase-like protein (DsbD) in *Escherichia coli*.
#cross-references M01D:95354659
#accession S57220
#status preliminary
#molecule-type DNA
#residues 462-540 ##label MIS

GENETICS
#gene dsbd; cycZ; Cuta2; dlpZ
#keywords inner membrane; redox-active disulfide; transmembrane protein
SUMMARY #length 565 #molecular-weight 61795 #checksum 2685

Query Match 58.3%; Score 63; DB 2; Length 565;
Best Local Similarity 53.8%; Pred. No. 4,47e-01;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

DB 373 PYELERVIGDVA 385
1-||| | | | |
QY 2 PPFLLSRNTGEVRT 14

RESULT 8
ENTRY S50482 #type complete
TITLE hypothetical protein YER024w - yeast (*Saccharomyces cerevisiae*)
ORGANISM #formal_name *Saccharomyces cerevisiae*
DATE 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 21-Nov-1997
S50482
#cross-references EMBL:U018778; NID:9603592; PID:9603616; MIPS:YER024w
REFERENCE
#authors Dietrich, F.S.
#submission submitted to the EMBL Data Library, December 1994
#description The sequence of *S. cerevisiae* cosmid 9537, 9581, 9495, 9867, and lambda clone 5898.
#accession S50482
#molecule-type DNA
#residues 1-923 ##label DIE
#cross-references EMBL:U018778; NID:9603592; PID:9603616; MIPS:YER024w
GENETICS
#map_position 5R
SUMMARY #length 923 #molecular-weight 10333 #checksum 767

Query Match 58.3%; Score 63; DB 2; Length 923;
Best Local Similarity 53.8%; Pred. No. 4,47e-01;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 722 MFMISWVIGSIRS 734
 QY 3 MFLSRTNGEVRT 15

RESULT 9
 ENTRY
 TITLE
 ALTERNATE_NAMES
 ORGANISM
 DATE

ASPMN2 #type complete
 asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 2 -
 garden pea
 asparagine synthetase (glutamine-hydrolyzing)
 #formal_name Pisum sativum #common_name garden pea
 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
 05-Sep-1997

ACCESSIONS
 REFERENCE
 #authors
 #journal
 #title

S11443
 S11443
 Tsai, F.Y.; Coruzzi, G.M.
 EMBO J. (1990) 9:323-332
 Dark-induced and organ-specific expression of two asparagine
 synthetase genes in Pisum sativum.
 #cross-references M01D:90151604
 #accession S11443
 ##molecule_type DNA
 ##residues 1-583 ##label TSA
 ##cross-references EMBL:X52180; NID:g20651; PID:g20652

GENETICS
 #gene
 CLASSIFICATION
 KEYWORDS
 FEATURE

AS2
 #superfamily asparagine synthase (glutamine-hydrolyzing)
 asparagine biosynthesis; ligase
 #active_site Cys (amide transfer) #status predicted
 #length 583 #molecular_weight 65649 #checksum 2836

SUMMARY
 2

Query Match
 Best Local Similarity 77.8%; Score 61; DB 1; Length 583;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPFMFSRK 330
 QY 1 TPFMFSRN 9

RESULT 10
 ENTRY
 TITLE
 ORGANISM
 DATE

S69182 #type complete
 asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 1 -
 Lotus japonicus
 #formal_name Lotus japonicus
 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
 08-Sep-1997

ACCESSIONS
 REFERENCE
 #authors
 #journal
 #title

S69182
 S69182
 Waterhouse, R.N.; Smyth, A.J.; Massonneau, A.; Prosser, I.M.;
 Clarkson, D.T.
 Plant Mol. Biol. (1996) 30:883-897
 Molecular cloning and characterization of asparagine
 synthetase from Lotus japonicus: dynamics of asparagine
 synthesis in N-sufficient conditions.
 #cross-references M01D:96270368
 #accession S69182
 ##status nucleic acid sequence not shown
 ##molecule_type mRNA
 ##residues 1-586 ##label WAT
 ##cross-references EMBL:X89409; NID:g897770; PID:g897771
 ##experimental_source strain B-129

GENETICS
 #gene
 CLASSIFICATION
 KEYWORDS
 SUMMARY

AS1
 #superfamily asparagine synthase (glutamine-hydrolyzing)
 asparagine biosynthesis; ligase
 #length 586 #molecular_weight 66461 #checksum 3372

Query Match
 Best Local Similarity 77.8%; Score 61; DB 2; Length 586;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPFMFSRK 330
 QY 1 TPFMFSRN 9

RESULT 11
 ENTRY
 TITLE
 ALTERNATE_NAMES
 ORGANISM
 DATE

ASPMN1 #type complete
 asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 1 -
 garden pea
 asparagine synthetase (glutamine-hydrolyzing)
 #formal_name Pisum sativum #common_name garden pea
 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
 05-Sep-1997

ACCESSIONS
 REFERENCE
 #authors
 #journal
 #title

S11444
 S11443
 Tsai, F.Y.; Coruzzi, G.M.
 EMBO J. (1990) 9:323-332
 Dark-induced and organ-specific expression of two asparagine
 synthetase genes in Pisum sativum.
 #cross-references M01D:90151604
 #accession S11444
 ##molecule_type DNA
 ##residues 1-586 ##label TSA
 ##cross-references EMBL:X52179; NID:g20649; PID:g20650

COMMENT
 This protein is one of a family of glutamine amidotransferases that
 have dual specificity to utilize either glutamine or ammonia as a
 substrate. Two functional domains have been identified: an
 aminator domain, which catalyzes the ammonia-dependent reaction,
 and a glutamine amide transfer domain, which binds glutamine and
 transfers the amide to the aminator domain.

GENETICS
 #gene
 CLASSIFICATION
 KEYWORDS
 FEATURE

AS1
 #superfamily asparagine synthase (glutamine-hydrolyzing)
 asparagine biosynthesis; ligase
 #domain aminator #label AMN
 #active_site Cys (amide transfer) #status predicted
 #length 586 #molecular_weight 66353 #checksum 3350

SUMMARY
 2

Query Match
 Best Local Similarity 77.8%; Score 61; DB 1; Length 586;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 323 TPFMFSRK 331
 QY 1 TPFMFSRN 9

RESULT 12
 ENTRY
 TITLE
 ORGANISM
 DATE

S52387 #type complete
 asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) -
 wild cabbage
 #formal_name Brassica oleracea #common_name wild cabbage
 08-May-1995 #sequence_revision 03-Nov-1995 #text_change
 17-Jul-1998

ACCESSIONS
 REFERENCE
 #authors
 #journal
 #title

S52387
 S52387
 Downs, C.G.; Pogson, B.J.; Davies, K.M.; Almitra, E.C.
 submitted to the EMBL Data Library, February 1995
 An asparagine synthetase cDNA clone from Broccoli (Brassica
 oleracea L.).
 #accession S52387
 ##molecule_type mRNA
 ##residues 1-586 ##label DOW
 ##cross-references EMBL:X84448; NID:g669056; PID:g669057

CLASSIFICATION
 KEYWORDS
 FEATURE

AS1
 #superfamily asparagine synthase (glutamine-hydrolyzing)
 asparagine biosynthesis; ligase
 #active_site Cys (amide transfer) #status predicted
 #length 586 #molecular_weight 65672 #checksum 3781

SUMMARY
 2

Query Match
 Best Local Similarity 77.8%; Score 61; DB 2; Length 586;

Best Local Similarity 77.8%; Pred. No. 1.09e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 323 TPFMLSRK 331
|||||:
QY 1 TPFMLSRN 9

RESULT 13
ENTRY T02978 #type complete
TITLE asparagine synthase (EC 6.3.1.1) - maize
ALTERNATE_NAMES asparagine synthetase
ORGANISM #formal_name Zea mays #common_name maize
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999

ACCESSIONS
REFERENCE T02978
#authors Chevalier, C.; Bourgeois, E.; Just, D.; Raymond, P.
#journal Plant J. (1996) 9:111
#title Metabolic regulation of asparagine synthetase gene expression in maize (Zeamays L.) root tips.

GENETICS
#accession T02978
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-586 #label CHE
#cross-references EMBL:X82849; NID:g984261; PID:g984262
#experimental_source cultivar var. DEA; root meristem

KEYWORDS
SUMMARY AS
ligase
#length 586 #molecular-weight 66577 #checksum 5815

Query Match 56.5%; Score 61; DB 2; Length 586;
Best Local Similarity 77.8%; Pred. No. 1.09e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPFMLSRK 330
|||||:
QY 1 TPFMLSRN 9

RESULT 14
ENTRY S69183 #type complete
TITLE asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 2 - Lotus japonicus
ORGANISM #formal_name Lotus japonicus
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 08-Sep-1997
ACCESSIONS S69183; S57932
REFERENCE S69183
#authors Waterhouse, R.N.; Smyth, A.J.; Massonneau, A.; Prosser, I.M.; Clarkson, D.T.
#journal Plant Mol. Biol. (1996) 30:883-897
#title Molecular cloning and characterisation of asparagine synthetase from Lotus japonicus: dynamics of asparagine synthesis in N-sufficient conditions.

GENETICS
#cross-references MUID:96270368
#accession S69183
#status nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-586 #label WAT
#cross-references EMBL:X89410; NID:g897772; PID:g897773
#experimental_source strain B-129

KEYWORDS
SUMMARY AS2
#superfamily asparagine synthase (glutamine-hydrolyzing) asparagine biosynthesis; ligase
#length 586 #molecular-weight 65969 #checksum 518

Query Match 56.5%; Score 61; DB 2; Length 586;
Best Local Similarity 77.8%; Pred. No. 1.09e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPFMLSRK 330
|||||:
QY 1 TPFMLSRN 9

RESULT 15
ENTRY T03602 #type complete
TITLE probable asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) - rice
ORGANISM #formal_name Oryza sativa #common_name rice
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999

ACCESSIONS
REFERENCE T03602
#authors Yamaya, T.; Higuchi, T.
#submission submitted to the EMBL Data Library, February 1996
#accession T03602

GENETICS
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-591 #label YAM
#cross-references EMBL:D83378; NID:d1095641; PID:d1019692
#experimental_source cv. Nipponbare

KEYWORDS
SUMMARY asparagine biosynthesis; ligase
#length 591 #molecular-weight 66227 #checksum 3113

Query Match 56.5%; Score 61; DB 2; Length 591;
Best Local Similarity 77.8%; Pred. No. 1.09e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 322 TPFMLSRK 330
|||||:
QY 1 TPFMLSRN 9

Search completed: Fri Jun 11 17:33:46 1999.
Job time : 94 secs.

KW CELL ADHESION: SIGNAL; TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
 KW CALCIUM-BINDING; REPEAT.
 FT SIGNAL 1 23
 FT PROPER 24 49
 FT CHAIN 50 99
 FT DOMAIN 50 615
 FT TRANSMEM 616 640
 FT DOMAIN 641 999
 FT REPEAT 50 158
 FT REPEAT 159 268
 FT REPEAT 269 383
 FT REPEAT 386 499
 FT REPEAT 910 935
 FT REPEAT 936 966
 FT CARBOHYD 110 110
 FT CARBOHYD 180 180
 FT CARBOHYD 459 459
 FT CARBOHYD 545 545
 SO SEQUENCE 999 AA; 107503 MW; 4891F6AE CRC32;
 Query Match 100.0%; Score 108; DB 1; Length 999;
 Best Local Similarity 100.0%; Pred. No. 4.34e-13;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 206 TPEFLSRNTEGVRT 220
 QY 1 TPEFLSRNTEGVRT 15
 RESULT 2
 ID DSGI HUMAN STANDARD; PRT; 1049 AA.
 AC 002413;
 DT 01-OCT-1993 (REL. 27, CREATED)
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DGI).
 GN DSGI.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATHARTINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KERATINOCYTES;
 RX MEDLINE: 91271279.
 RA WHEELER G.N., PARKER A.E., THOMAS C.L., ATALOTIS P., POYTER D.,
 RA ARNEAAN J., RUTMAN A.J., PIDDSLEY S.C., WATT F.M., REES D.A.,
 RA BUXTON R.S., NAGEE A.I.;
 RT "Desmosomal glycoprotein DGI, a component of intercellular desmosome
 junctions, is related to the cadherin family of cell adhesion
 molecules.";
 RT PROC. NATL. ACAD. SCI. U.S.A. 88:4796-4800(1991).
 RL -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.
 INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
 FILAMENTS MEDIATING CELL-CELL ADHESION.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, PENIS AND OSOPHAGUS.
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE
 DESMOsome SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X56654; G30506; -;
 DR PIR: S16906; IJHDG1.
 DR MIM: 125670; -;
 DR PROSITE: PS00232; CADHERIN; 2.

DR PFAM: PF00028; cadherin; 4.
 DR HSSP: P09803; 1EDH.
 KW CELL ADHESION: SIGNAL; TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
 KW CALCIUM-BINDING; REPEAT.
 FT SIGNAL 1 23
 FT PROPER 24 49
 FT CHAIN 50 1049
 FT DOMAIN 50 545
 FT TRANSMEM 546 570
 FT DOMAIN 571 1049
 FT REPEAT 571 1049
 FT REPEAT 159 270
 FT REPEAT 271 385
 FT REPEAT 386 497
 FT REPEAT 813 839
 FT REPEAT 840 869
 FT REPEAT 870 899
 FT REPEAT 900 927
 FT REPEAT 928 956
 FT DOMAIN 969 1019
 FT CARBOHYD 36 36
 FT CARBOHYD 110 110
 FT CARBOHYD 180 180
 SO SEQUENCE 1049 AA; 113715 MW; FDD79961 CRC32;
 Query Match 81.5%; Score 88; DB 1; Length 1049;
 Best Local Similarity 66.7%; Pred. No. 1.08e-07;
 Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 206 SPFLINRNTGEIRT 220
 QY 1 TPEFLSRNTEGVRT 15
 RESULT 3
 ID DSGI BOVIN STANDARD; PRT; 1043 AA.
 AC 003763;
 DT 01-OCT-1993 (REL. 27, CREATED)
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DGI).
 GN DSGI.
 OS BOS TAURUS (BOVINE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-MUZZLE EPITHELIUM;
 RA KOCH P.J., GOLDSCHMIDT M.D., ZIMBELMANN R., FRANK W.W.;
 RA SUBMITTED (MAR-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
 RL [2]
 RP SEQUENCE OF 44-1043 FROM N.A.
 RC TISSUE-MUZZLE EPITHELIUM;
 RX MEDLINE: 91168965.
 RA KOCH P.J., WALSH M.J., SCHWELZ M., GOLDSCHMIDT M.D.,
 RA ZIMBELMANN R., FRANK W.W.;
 RT "Identification of desmoglein, a constitutive desmosomal
 glycoprotein, as a member of the cadherin family of cell adhesion
 molecules.";
 RT EUR. J. CELL BIOL. 53:1-12(1990).
 RL [3]
 RN REVISIONS, AND SEQUENCE OF 101-123.
 RP MEDLINE: 92037656.
 RX KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R.,
 RA FRANK W.W.;
 RT "Complete amino acid sequence of the epidermal desmoglein precursor
 polypeptide and identification of a second type of desmoglein gene.";
 RT EUR. J. CELL BIOL. 55:200-208(1991).
 RL [4]
 RN SEQUENCE OF 44-493 FROM N.A.
 RP MEDLINE: 91097553.
 RX GOODWIN L., HILL J.E., RAYNOR K., RASZI L., MANABE M., COMIN P.;
 RA "Desmoglein shows extensive homology to the cadherin family of cell

```

RT adhesion molecules".
RL BIOCHEM. BIOPHYS. RES. COMMUN. 173:1224-1230(1990).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EPIDERMIS, MZZLE, TONGUE AND OESOPHAGUS.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE
CC DESMOsome SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X58466; G307;
DR EMBL; X57784; G436062;
DR EMBL; M58165; G552318;
DR PIR; S14603; IJB01.
DR PROSITE; PS00232; CADHERIN; 2.
DR PFAM; PF00028; cadherin; 3.
DR HSP; P09803; IEDH.
DR CELL ADHESION; SIGNAL; TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
KW CALCIUM-BINDING; REPEAT.
FT SIGNAL 1 23
FT PROPEP 24 49
FT CHAIN 50 1043
FT DOMAIN 50 548
FT TRANSMEM 549 573
FT DOMAIN 574 1043
FT REPEAT 50 158
FT REPEAT 159 270
FT REPEAT 271 385
FT REPEAT 386 498
FT REPEAT 499 845
FT REPEAT 846 875
FT REPEAT 876 905
FT REPEAT 906 933
FT REPEAT 934 962
FT DOMAIN 963 1012
FT CARBOHYD 110 110
FT CARBOHYD 180 180
FT CARBOHYD 496 496
FT CONFLICT 124 124
SQ SEQUENCE 1043 AA; 112243 MW; 13898584 CRC32;

Query Match
Best Local Similarity 75.0%; Score 81; DB 1; Length 1043;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 206 SPANINRYTGEIRT 220
OY 1 TPELLSRNTGEVRT 15

```

```

RP SEQUENCE FROM N.A.
RL DANG V.D., BOLOTIN-FUKUHARA M., DAIGNAN-FORMIER B.;
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DDO DATA BANKS.
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN-5288C / AB972;
RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,
RA FAVELLO A., FULTON L., GATUNG S., GRECO T., KIRSTEN J., KUCABA T.,
RA HALLSWORTH K., HAWKINS J., HILLIER L., JIER M., JOHNSON D.,
RA JOHNSTON L., LANGSTON Y., LATREILLE P., LE T., MARDIS E., MENEZES S.,
RA MILLER N., NHAN M., PAULEY A., PELUSO D., RIFKEN L., RILES L.,
RA TATOH A., TREVASIS E., VIGNATI D., WILCOX L., WOLDMAN P., VAUDIN M.,
RA WILSON R., WATERSON R., EMBL/GENBANK/DDO DATA BANKS.
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DDO DATA BANKS.
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE = AMP +
CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC AMIDOTRANSFERASES.
CC -----
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z48675; G747902;
DR EMBL; U40829; G1066479;
DR SGD; L0002732; ASN1.
DR PROSITE; PS00443; GATASE-TYPE-1; 1.
DR PFAM; PF00310; GATase.2; 1.
DR PFAM; PF00733; Asn-synthase; 1.
KW LIGASE; ASPARAGINE BIOSYNTHESIS; GLUTAMINE AMIDOTRANSFERASE;
KW MULTIGENE FAMILY.
FT INIT_MER 0 0
FT ACT_SITE 1 1
FT ACT_SITE 1 1
SQ SEQUENCE 571 AA; 64339 MW; 6FA2535B CRC32;

Query Match
Best Local Similarity 60.2%; Score 65; DB 1; Length 571;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 346 TPELLSRN 354
OY 1 TPELLSRN 9

```

```

RESULT 4
ID ASN1-YEAST STANDARD; PRT; 571 AA.
AC P49089;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] 1 (EC 6.3.5.4)
DE (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE 1).
GN ASN1 OR YPR145W OR P9659.3.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCYCETES; SACCCHAROMYCETALES;
OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.
RN [1]

```

```

RESULT 5
ID ASN2-YEAST STANDARD; PRT; 571 AA.
AC P49090;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] 2 (EC 6.3.5.4)
DE (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE 2).
GN ASN2 OR YGR124W OR G6358.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCYCETES; SACCCHAROMYCETALES;
OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-5288C / FY1679;
RX MEDLINE; 97197982.
RA VAN DYCK L., TETTELIN H., PURNELLE B., GOFFEAU A.;
RT "An 18.3 kb DNA fragment from yeast chromosome VII carries four
RT unknown open reading frames", the gene for an Asn synthase, remnants
RT of Ty and three tRNA genes";
RL YEAST 13:171-176(1997).
RN [2]
RP SEQUENCE FROM N.A.

```

RA VAN DYCK L., SKALA J., DE WERGFOSSE P., PURNELLE B., TALIA E.,
 RA NAMROCKI A., DEL BINO S., GOFFEAU A.;
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARATE + L-GLUTAMINE -> AMP +
 CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
 CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
 CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
 CC AMIDOTRANSFERASES.
 CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X83099; G642347; -;
 DR EMBL: 272909; E243463; -;
 DR SGD: L0003156; ASN2.
 DR PROSITE: PS00443; GATASE_TYPE_II; 1.
 DR PFAM: PF00310; GATASE_2; 1.
 DR PFAM: PF00733; Asn_synthase; 1.
 KW LIGASE; ASPARAGINE BIOSYNTHESIS; GLUTAMINE AMIDOTRANSFERASE;
 KM MULTIGENE FAMILY.
 FT INIT MET 0 0 BY SIMILARITY.
 FT ACT_SITE 1 1 GATASE (BY SIMILARITY).
 SQ SEQUENCE 571 AA; 64461 MW; 92609736 CRC32;

Query Match 60.2%; Score 65; DB 1; Length 571;
 Best Local Similarity 88.9%; Pred. No. 4,12e-02;
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 345 TPMEFLLSRK 353
 QY 1 TPMEFLLSRN 9

RESULT 6
 ID DSD, ECOLI STANDARD; PRT; 488 AA.
 AC P36655; P76796;
 DT 01-JUN-1994 (REL. 29, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE THIOL-DISULFIDE INTERCHANGE PROTEIN DSD (C-TYPE CYTOCHROME
 DE BIOGENESIS PROTEIN CYCZ) (INNER MEMBRANE COPPER TOLERANCE PROTEIN).
 GN DSD OR CYCZ OR COTAZ OR DIPZ.
 OS ESCHERICHIA COLI.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
 OC ESCHERICHIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 95349398.
 RA CROOKE H., COLE J.;
 RT "The biogenesis of c-type cytochromes in Escherichia coli requires a
 RT membrane-bound protein Dipz, with a protein disulfide
 RT isomerase-like domain";
 RL MOL. MICROBIOL. 15:1139-1150(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / W3110;
 RX MEDLINE: 95349397.
 RA FONG S.-T., CAMAKARIS J., LEE B.T.O.;
 RT "Molecular genetics of a chromosomal locus involved in copper
 RT tolerance in Escherichia coli K-12";
 RL MOL. MICROBIOL. 15:1127-1137(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 95343462.
 RA BURLAND V.D., PLUNKETT G. III, SOTIA H.J., DANIELS D.L.,

RA BLATTNER F.R.;
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
 RT region from 92.8 through 100 minutes";
 RL NUCLEIC ACIDS RES. 23:2105-2119(1995).
 RN [4]
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
 RX MEDLINE: 95354659.
 RA MISSIRAKIS D., SCHWAGER F., RAINA S.;
 RT "Identification and characterization of a new disulfide
 RT isomerase-like protein (Dsd) in Escherichia coli";
 RL EMBL J. 14:3415-3424(1995).
 RN [5]
 RP SEQUENCE OF 1-5.
 RA MISSIRAKIS D., HUGHES G.J., FRUTIGER S., PAQUET N., RAINA S.;
 RL SUBMITTED (MAY-1995) TO THE SWISS-PROT DATA BANK.
 RN [6]
 RP CHARACTERIZATION.
 RX MEDLINE: 95046265.
 RA SAMBONGI Y., FERGUSON S.J.;
 RT "Specific thiol compounds complement deficiency in c-type cytochrome
 RT biogenesis in Escherichia coli carrying a mutation in a
 RT membrane-bound disulfide isomerase-like protein";
 RL FEBS LETT. 335:235-238(1994).
 CC -1- FUNCTION: REQUIRED FOR DISULFIDE BOND FORMATION IN SOME
 CC PERIPLASMIC PROTEINS. ONE OF ITS FUNCTION COULD BE TO MAINTAIN
 CC THE CYS RESIDUES IN APOCYTOCHROMES C IN A REDUCED STATE TO ALLOW
 CC COVALENT LINKAGE WITH THE HEME.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X77707; G871029; -;
 DR EMBL: U14003; G536980; ALT_INIT.
 DR EMBL: AE000486; G1790578; ALT_INIT.
 DR EMBL: Z36905; G581055; -;
 DR PIR: S42064; S42064.
 DR ECOGENE: EG12178; DSD.
 DR PROSITE: PS00194; THIOREDOXIN; 1.
 DR HSP: P10399; ITRU.
 KW REDOX-ACTIVE CENTER; TRANSMEMBRANE; INNER MEMBRANE;
 KW CYTOCHROME C-TYPE BIOGENESIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 95349398.
 RA CROOKE H., COLE J.;
 RT "The biogenesis of c-type cytochromes in Escherichia coli requires a
 RT membrane-bound protein Dipz, with a protein disulfide
 RT isomerase-like domain";
 RL MOL. MICROBIOL. 15:1139-1150(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / W3110;
 RX MEDLINE: 95349397.
 RA FONG S.-T., CAMAKARIS J., LEE B.T.O.;
 RT "Molecular genetics of a chromosomal locus involved in copper
 RT tolerance in Escherichia coli K-12";
 RL MOL. MICROBIOL. 15:1127-1137(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 95343462.
 RA BURLAND V.D., PLUNKETT G. III, SOTIA H.J., DANIELS D.L.,

Query Match 58.3%; Score 63; DB 1; Length 488;
 Best Local Similarity 53.8%; Pred. No. 1,14e-01;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 296 PVFLERVIGDWM 308

QY 2 MFLLSRNTGEVR 14
 |||||
 RESULT 7
 ID YEEL YEAST STANDARD: PRT: 923 AA.
 AC P40017
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE HYPOTHEICAL 103.3 KD PROTEIN IN PRO3-GCD11 INTERGENIC REGION.
 GN YER024W
 OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST)
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCAROMYCETALES;
 CC SACCAROMYCETACEAE; SACCAROMYCES.
 CC [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972.
 RA DIETRICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,
 RA AVILES E., BERNI A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,
 RA CHUNG E., DUNCAN M., GUZMAN E., HARTZEL G., HUNICKE-SMITH S.,
 RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,
 RA MOSKALC D., NAKAHARA K., NAMATH A., NORGREN R., OEBNER P., OH C.,
 RA PEREL F.X., ROBERTS D., SEHL P., SCHRAMM S., SHOGREN T., SMITH V.,
 RA TAYLOR P., WEI Y., YELTON M., BOTSTEIN D., DAVIS R.W.;
 RA SUBMITTED (DEC-1994) TO EMBL/GENBANK/DDJ DATA BANKS.
 CC [1] SIMILARITY: BELONGS TO THE CARNITINE/CHOLINE ACETYLTRANSFERASE
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DB 722 MFLLSRNTGEVR 734
 |||||
 QY 3 MFLLSRNTGEVR 15
 |||||
 Query Match 58.3%; Score 63; DB 1; Length 923;
 Best Local Similarity 53.8%; Pred. No. 1,14e-01;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 DB 722 MFLLSRNTGEVR 734
 |||||
 QY 3 MFLLSRNTGEVR 15
 |||||
 RESULT 8
 ID ASNS SANAU STANDARD: PRT: 524 AA.
 AC 024338;
 DT 15-DEC-1998 (REL. 37, CREATED)
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE ASPARAGINE SYNTHETASE (GLUTAMINE-HYDROLYZING) (EC 6.3.5.4) (GLUTAMINE-
 DE DEPENDENT ASPARAGINE SYNTHETASE).
 GN ANDL.
 OS SANDERSONIA AURANTIACA (CHRISTMAS-BELLS) (CHINESE-LANTERN LILY).
 OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 CC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; LILIALES;
 CC LILIACEAE; SANDERSONIA.
 CC [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TERAL;
 RA EASON J.R., KING G.A.;
 RA "Nucleotide sequence of cDNA encoding asparagine synthetase from
 RT *Sanderonia aurantiaca*.";
 RT (IN) PLANT GENE REGISTER, PGR97-112.

CC -1- FUNCTION: COTID PLAY A ROLE IN REMOBILIZATION OF NITROGEN IN
 CC FLOWERS DURING SENESENCE.
 CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE -> AMP +
 CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
 CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
 CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
 CC AMIDOTRANSFERASES.
 CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DB 321 TPEFLMSRK 329
 |||||
 QY 1 TPEFLMSRK 9
 |||||
 Query Match 56.5%; Score 61; DB 1; Length 524;
 Best Local Similarity 77.8%; Pred. No. 3.06e-01;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 DB 321 TPEFLMSRK 329
 |||||
 QY 1 TPEFLMSRK 9
 |||||
 RESULT 9
 ID ASN2-PEA STANDARD: PRT: 582 AA.
 AC P19252;
 DT 01-NOV-1990 (REL. 16, CREATED)
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
 DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
 DE ASPARAGINE SYNTHETASE, ROOT (GLUTAMINE-HYDROLYZING) (EC 6.3.5.4)
 DE (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE).
 GN AS2.
 OS PISUM SATIVUM (GARDEN PEA).
 OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 CC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 CC FABACEAE; PAPILIONOIDEAE; PISUM.
 CC [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. SPARKLE; TISSUE=ROOT;
 RX MEDLINE: 90151604.
 RA TSAI F.Y., CORUZZI G.M.;
 RA "Dark-induced and organ-specific expression of two asparagine
 RT synthetase genes in *Pisum sativum*.";
 RL EMBO J. 9:323-332(1990).
 CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE -> AMP +
 CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
 CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
 CC -1- TISSUE SPECIFICITY: ROOTS.
 CC -1- INDUCTION: DARK-INDUCED.
 CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
 CC AMIDOTRANSFERASES.
 CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: X52180; G20652; -
DR PIR: S11443; AJPAN2.
DR PROSITE: PS00443; GATASE_TYPE_II; 1.
DR PFAM: PF00310; GATase_2; 1.
DR PFAM: PF00733; Asn_synthase; 1.
KW LIGASE: ASPARAGINE BIOSYNTHESIS; GLUTAMINE AMIDOTRANSFERASE;
KW MULTIGENE FAMILY.
FT INIT MET 0 BY SIMILARITY.
FT ACT SITE 1 GATASE (BY SIMILARITY).
SQ SEQUENCE 582 AA: 65518 MM; 793421FA CRC32;
Query Match 56.5%; Score 61; DB 1; Length 582;
Best Local Similarity 77.8%; Pred. No. 3,06e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
DB 321 TPEFLMSRK 329
OY 1 TPEFLMSRN 9
|||||:

RESULT 10 STANDARD; PRT; 583 AA.
ID ASNS_ARATH
AC P49078;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE).
DE DEPENDENT ASPARAGINE SYNTHETASE.
GN ASN1.
UN ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
CC EUPHRYOTA: VIRIDIPHYTES: STREPTOPHYTA: EMBRYOPHYTA: TRACHEOPHYTA;
CC EUPHYLOPHITES: SPERMATOPHYTA: MAGNOLIOPHYTA: EUDICOTYLEDONS; ROSIDAE;
OC CAPRALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:CV. COLUMBIA;
RX MEDLINE: 95148732.
RA LAM H.M., PENG S.S., CORUZZI G.M.;
RT "Metabolic regulation of the gene encoding glutamine-dependent
RT asparagine synthetase in Arabidopsis thaliana."
RL PLANT PHYSIOL. 106:1347-1357(1994).
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE -> AMP +
CC PYROPHOSPHATE + L-ASPARGINE + L-GLUTAMATE.
CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC AMIDOTRANSFERASES.
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L29083; G507946; -
DR PROSITE: PS00443; GATASE_TYPE_II; 1.
DR PFAM: PF00310; GATase_2; 1.
DR PFAM: PF00733; Asn_synthase; 1.
DR HSP: P17169; IGMS.
KW LIGASE: ASPARAGINE BIOSYNTHESIS; GLUTAMINE AMIDOTRANSFERASE;
KW MULTIGENE FAMILY.
FT INIT MET 0 BY SIMILARITY.
FT ACT SITE 1 GATASE (BY SIMILARITY).
SQ SEQUENCE 583 AA: 65489 MM; F8FC9672 CRC32;
Query Match 56.5%; Score 61; DB 1; Length 583;
Best Local Similarity 77.8%; Pred. No. 3,06e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
DB 321 TPEFLMSRK 329
OY 1 TPEFLMSRN 9
|||||:

OY 1 TPEFLMSRN 9
|||||:
RESULT 11 STANDARD; PRT; 585 AA.
ID ASNS_TRIYS
AC O24661;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE).
GN AS.
UN TRIPHSARIA VERSICOLOR.
CC EUPHRYOTA: VIRIDIPHYTES: STREPTOPHYTA: EMBRYOPHYTA;
CC TRACHEOPHYTA: EUPHYLOPHITES: SPERMATOPHYTA: MAGNOLIOPHYTA;
OC EUDICOTYLEDONS; ASTERIDAE; GENTIANACEAE; LAMIALES; SCROPHULARIACEAE;
OC TRIPHSARIA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE:ROOT.
RA DELAVALL P., ESTABROOK E., ALBRECHT H., WOBEL R., YODER J.;
RT "Host root exudate increases expression of asparagine synthetase in
RT the roots of a hemiparasitic plant Triphysaria (Scrophulariaceae).";
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE -> AMP +
CC PYROPHOSPHATE + L-ASPARGINE + L-GLUTAMATE.
CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC AMIDOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF014055; G2439280; -
DR EMBL: AF014056; G2439282; -
DR EMBL: AF014057; G2439284; -
DR PROSITE: PS00443; GATASE_TYPE_II; 1.
DR PFAM: PF00310; GATase_2; 1.
DR PFAM: PF00733; Asn_synthase; 1.
KW LIGASE: ASPARAGINE BIOSYNTHESIS; GLUTAMINE AMIDOTRANSFERASE;
KW MULTIGENE FAMILY.
FT INIT MET 0 BY SIMILARITY.
FT ACT SITE 1 GATASE (BY SIMILARITY).
SQ SEQUENCE 585 AA: 65560 MM; B3ACC66E CRC32;
Query Match 56.5%; Score 61; DB 1; Length 585;
Best Local Similarity 77.8%; Pred. No. 3,06e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
DB 321 TPEFLMSRK 329
OY 1 TPEFLMSRN 9
|||||:
RESULT 12 STANDARD; PRT; 585 AA.
ID ASNS_BRAOL
AC P49091;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE).
DE BRASSICA OLERACEA (CAULIFLOWER).
OS EUPHYLOPHITES: VIRIDIPHYTES: STREPTOPHYTA: EMBRYOPHYTA: TRACHEOPHYTA;
OC EUPHYLOPHITES: SPERMATOPHYTA: MAGNOLIOPHYTA: EUDICOTYLEDONS; ROSIDAE;
OC CAPRALES; BRASSICACEAE; BRASSICA.
RN [1]
RP SEQUENCE FROM N.A.

```

CC -----
DR EMBL: X82849; G984262; -.
DR DR MATZEDB; 79071; GATASE_TYPE.IT; 1.
DR PROSITE; PS00443; GATase_2; 1.
DR PFAM; PF00310; GATase_2; 1.
DR PFAM; PF00733; Asn_synthase; 1.
DR KJ LIGASE; ASPARAGINE BIOSYNTHESIS; GLUTAMINE AMIDOTRANSFERASE.
DR INT MET 0 BY SIMILARITY.
FT ACT_SITE 1 GATASE (BY SIMILARITY).
SQ SEQUENCE 585 AA; 66446 MW; B24B0478 CRC32;

Query Match
Best Local Similarity 77.8%; Score 61; DB 1; Length 585;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 321 TPFMTSRK 329
0Y 1 TPFMLSRN 9

RESULT 14
AC ASN2_LOTJA STANDARD; PRT; 585 AA.
AC P49093;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDOLYZIN] 2 (EC 6.3.5.4)
DE (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE 2).
GN AS2.
OS LOTUS JAPONICUS.
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC FABALES; FABACEAE; PAPILIONOIDEAE; LOTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV, GIRU / B-129;
RX MEDLINE; 96270368.
RA WATERHOUSE R.N., SMYTH A.J., MASSONEAU A., PROSSER I.M.,
RA CLARKSON D.T.;
RT "Molecular cloning and characterisation of asparagine synthetase from
RT Lotus japonicus: dynamics of asparagine synthesis in N-sufficient
RT conditions.";
RL PLANT MOL. BIOL. 30:883-897(1996)
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARATE + L-GLUTAMINE = AMP +
CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC AMIDOTRANSFERASES.
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isd-sb.ch/announce
CC or send an email to license@isd-sb.ch).
CC CC
DR EMBL; X89410; G897773; -.
DR PROSITE; PS00443; GATASE_TYPE.IT; 1.
DR PFAM; PF00310; GATase_2; 1.
DR PFAM; PF00733; Asn_synthase; 1.
DR HSPB; P17169; IGMS.
DR KJ LIGASE; ASPARAGINE BIOSYNTHESIS; GLUTAMINE AMIDOTRANSFERASE;
KW MULTIGENE FAMILY.
FT INIT_MET 0 BY SIMILARITY.
FT ACT_SITE 1 GATASE (BY SIMILARITY).
SQ SEQUENCE 585 AA; 65838 MW; 2FE40574 CRC32;

Query Match
Best Local Similarity 77.8%; Score 61; DB 1; Length 585;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

DB 321 TPMFLMSRK 329
 111111111:
 QY 1 TPMFLLSRN 9

RESULT 15
 ASNL_PEA STANDARD: PRT: 585 AA.

AC P19251;
 DT 01-NOV-1990 (REL. 16, CREATED)
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE ASPARAGINE SYNTHETASE, NODULE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4)
 DE (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE).
 GN ASI.
 OS PISUM SATIVUM (GARDEN PEA).
 OC EUKARYOTA; VIRIDIPHYTES; EMBRYOPHYTES; TRACHEOPHYTES;
 OC EUPHYLOPHYTES; SPERMATOPHYTES; MAGNOLIOPHYTES; EUDICOTYLEDONS; ROSIDAE;
 OC FABALES; FABACEAE; PAPILIONOIDEAE; PISUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV, SPARKLE; TISSUE-ROOT NODULES;
 RX MEDLINE; 90151604.
 RA TSAI F.Y., CORUZZI G.M.;
 RT "Dark-induced and organ-specific expression of two asparagine
 synthetase genes in Pisum sativum.";
 RL EMBO J. 9:323-332(1990).
 CC -1 CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE -> AMP +
 CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
 CC -1 PATHWAY: ASPARAGINE BIOSYNTHESIS.
 CC -1 TISSUE SPECIFICITY: NODULE.
 CC -1 INDUCTION: DARK-INDUCED.
 CC -1 SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
 CC AMIDOTRANSFERASES.
 CC -1 SIMILARITY: TO OTHER ASN SYNTHETASES.

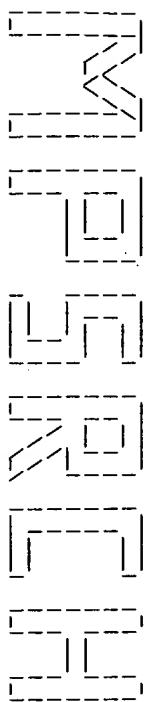
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

DR EMBL; X52179; G20650; -.
 DR PIR; S11444; AJPMN1.
 DR PROSITE; PS00443; GATASE_TYPE_1; 1.
 DR PIRAM; PF00310; GATase_2; 1.
 DR PIRAM; PF00733; Asn_synthase; 1.
 DR HSSP; P17169; IGMS.
 DR KW LIGASE; ASPARAGINE BIOSYNTHESIS; GLUTAMINE AMIDOTRANSFERASE;
 DR MULTIGENE FAMILY.
 DR INIT_MET 0
 DR ACT_SITE 1 1 BY SIMILARITY.
 DR GATASE (BY SIMILARITY).
 SO SEQUENCE 585 AA; 66222 MW; 9AB4FBC5 CRC32;

Query Match 56.5%; Score 61; DB 1; Length 585;
 Best Local Similarity 77.8%; Pred. No. 3.06e-01;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 322 TPMFLMSRK 330
 111111111:
 QY 1 TPMFLLSRN 9

Search completed: Fri Jun 11 17:29:30 1999
 Job time : 8 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 11 17:29:48.1999; MasPar time 6.28 Seconds

Tabular output not generated. 130.422 Million cell updates/sec

Title: >US-08-991-628-4
Description: (1-15) from US08991628.pep
Perfect Score: 108
Sequence: 1 TPFMLLSRNTGEVRT 15

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

sptrembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mmc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_protist 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 27.267; Variance 32.331; scale 0.843

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	90	83.3	993	11 035902	DESMOLEIN 3 (FRAGMENT	5.59e-08
2	64	59.3	336	5 009661	HYPOTHEICAL 37.7 KD P	1.22e-01
3	63	58.3	291	2 004824	ORF6 PROTEIN PRECURSOR	2.03e-01
4	62	57.4	322	5 020732	F53H10.1.	3.35e-01
5	61	56.5	557	3 042902	ASPARAGINE SYNTHETASE.	5.50e-01
6	61	56.5	581	10 P93167	ASPARAGINE SYNTHETASE.	5.50e-01
7	61	56.5	585	10 065329	ASPARAGINE SYNTHETASE.	5.50e-01
8	61	56.5	586	10 024483	ASPARAGINE SYNTHETASE.	5.50e-01
9	61	56.5	586	10 P93618	ASPARAGINE SYNTHETASE.	5.50e-01
10	61	56.5	586	10 040328	ASPARAGINE SYNTHETASE.	5.50e-01
11	59	54.6	148	1 058871	148AA LONG HYPOTHEICAL	1.46e+00
12	59	54.6	820	4 015039	KIAA0327 PROTEIN.	1.46e+00
13	58	53.7	593	14 010288	RIBONUCLEOTIDE REDUCTA	2.36e+00
14	57	52.8	483	10 049927	PS4 PROTEIN.	3.79e+00
15	56	51.9	300	2 069782	DARA.	6.05e+00
16	56	51.9	496	1 058705	496AA LONG HYPOTHEICAL	6.05e+00
17	56	51.9	579	10 P93168	ASPARAGINE SYNTHETASE	6.05e+00
18	55	50.9	366	2 050455	PUTATIVE MAGNESIUM AND	9.59e+00
19	55	50.9	501	13 P87358	ZNR-1.	9.59e+00

21	55	50.9	810	4 075279	KIAA0345-LIKE 3.	9.59e+00
22	54	50.0	163	2 084262	HYPOTHEICAL 18.9 KD P	1.51e+01
23	54	50.0	170	6 029035	CYCLOIN B (FRAGMENT).	1.51e+01
24	54	50.0	244	2 044500	CYCLOHOME BD.	1.51e+01
25	54	50.0	244	2 085222	ANABROBIC REGULATORY P	1.51e+01
26	54	50.0	309	5 001259	T20D3.8 PROTEIN.	1.51e+01
27	54	50.0	610	5 077330	MAL3P3.14 PROTEIN.	1.51e+01
28	54	50.0	762	2 P96793	XYLO.	1.51e+01
29	54	50.0	792	4 075278	KIAA0345-LIKE 2.	1.51e+01
30	54	50.0	1807	3 013661	HYPOTHEICAL 229.9KD P	1.51e+01
31	54	50.0	1919	3 042998	HYPOTHEICAL 217.7 KD	1.51e+01
32	53	49.1	215	5 044017	ERD2 GENE.	2.36e+01
33	53	49.1	319	2 086864	STRT.	2.36e+01
34	53	49.1	373	2 049979	CORA.	2.36e+01
35	53	49.1	393	2 008395	CITRATE SYNTHASE (EC 4	2.36e+01
36	53	49.1	433	14 041107	A623R PROTEIN.	2.36e+01
37	53	49.1	453	5 022370	T10B10.2 PROTEIN.	2.36e+01
38	53	49.1	578	2 P72723	POTENTIAL FMN-PROTEIN.	2.36e+01
39	53	49.1	736	2 024816	POLYPHOSPHATE KINASE.	2.36e+01
40	53	49.1	814	10 064681	F22013.35.	2.36e+01
41	53	49.1	1163	5 018820	SIMILARITY TO INSULIN	2.36e+01
42	53	49.1	1220	13 098864	PATCHED PROTEIN.	2.36e+01
43	53	49.1	1468	2 067762	RNA POLYMERASE BETA SU	2.36e+01
44	53	49.1	1607	5 094599	LCFACAS.	2.36e+01
45	52	48.1	326	5 018666	C47D12.3 PROTEIN.	3.67e+01

ALIGNMENTS

RESULT	ID	Sequence	Score	DB ID	Description	Pred. No.
1	035902	PRELIMINARY;	993 AA.			
AC	035902	PRELIMINARY;	993 AA.			
DT	01-JAN-1998	(TREMBLERL. 05, CREATED)				
DT	01-JAN-1998	(TREMBLERL. 05, LAST SEQUENCE UPDATE)				
DT	01-NOV-1998	(TREMBLERL. 08, LAST ANNOTATION UPDATE)				
DE	DESMOLEIN 3 (FRAGMENT).					
GN	DSG3.					
OS	MUS MUSCULUS (MOUSE).					
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;					
OC	SCIROGNATHI; MURIDAE; MURINAE; MUS.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-BALB/C.					
RA	ISHIKAWA H., LI K., UETTO J.,					
RL	SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.					
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).					
DR	EMBL; U86016; G2290200; -.					
DR	PROSITE; PS00232; CADHERIN; 2.					
DR	PFAM; PF00028; cadherin; 4.					
KW	CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.					
FT	NON_TER	993				
FT	SEQUENCE	993 AA; 107888 MW; 881794BD CRC32;				
SO	Query Match	83.3%; Score 90; DB 11; Length 993;				
	Best local Similarity	85.7%; Pred. No. 5.59e-08;				
	Matches	12; Conservative	2; Mismatches	0; Indels	0; Gaps	0;
DB	207	SPTFLSRNTGEVRT 220				
QY	2	TPFMLLSRNTGEVRT 15				
RESULT	2	PRELIMINARY;	PRT; 336 AA.			
AC	009661	PRELIMINARY;	PRT; 336 AA.			
DT	01-NOV-1996	(TREMBLERL. 01, CREATED)				
DT	01-NOV-1996	(TREMBLERL. 01, LAST SEQUENCE UPDATE)				
DT	01-NOV-1998	(TREMBLERL. 08, LAST ANNOTATION UPDATE)				
DE	HYPOTHEICAL 37.7 KD PROTEIN ZK1177.6 IN CHROMOSOME 11.					
GN	ZK1177.6.					
OS	CAENORHABDITIS ELEGANS.					
OC	EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;					
OC	EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;					
OC	RHABDITINA; RHABDITIDAE; RHABDITIDAE; RHABDITIDAE; CAENORHABDITIS.					

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA ANDERSON K.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: WEAK, TO YEAST PROTEIN CDC20.
DR EMBL: U21321; 6887854; -.
DR NCBI: U21321; 6887854; -.
DR WORMPEP: ZK177.6; CE02095.
DR PFAM: PF00400; G-beta; 1.
KW HYPOTHEICAL PROTEIN.
SQ SEQUENCE 336 AA; 37668 MW; 2DC77B26 CRC32;

Query Match
Best Local Similarity 59.3%; Score 64; DB 5; Length 336;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 187 PMFLSRHTGTVRT 200
    |||:|||||
OY 2 PMFLSRHTGTVRT 15

RESULT 3
ID 004824 PRELIMINARY; PRT; 291 AA.
AC 004824;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)
DE ORF6 PROTEIN PRECURSOR.
OS SALMONELLA TYPHIMURIO.
OC PLASMID 90 KB VIRULENCE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2;
KC MEDLINE: 93316852.
RA FRIEDRICH M.J., KINSEY N.E., VILLA J., KADNER R.J.;
RT "Nucleotide sequence of a 13.9 kb segment of the 90 kb virulence
   plasmid of Salmonella typhimurium: the presence of fimbrial
   biosynthetic genes."
RT MOL. MICROBIOL. 8:543-558(1993).
DR EMBL: L08613; G154246; -.
KW PLASMID; SIGNAL.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 291 ORF6 PROTEIN.
SQ SEQUENCE 291 AA; 30683 MW; C4D5ADDF CRC32;

Query Match
Best Local Similarity 58.3%; Score 63; DB 2; Length 291;
Matches 12; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Db 128 TLMFSLR-TGEVRT 141
    |||:|||||
OY 1 TLMFSLRNTGTVRT 15

RESULT 4
ID 020732 PRELIMINARY; PRT; 322 AA.
AC 020732;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)
DE F3H10.1
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA;
OC RHABDITIDA; RHABDITIDAE; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC WILKINSON J.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RX MEDLINE: 94150718.

```

```

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
RA SKALDON N., SMITH A., SONNHAMMER E., STADEN R., SUITON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPOAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
   elegans."
RL NATURE 368:32-38(1994).
DR EMBL: Z77664; E255929; -.
DR PFAM: PF00092; vwa; 1.
SQ SEQUENCE 322 AA; 36030 MW; 1C76BA97 CRC32;

Query Match
Best Local Similarity 57.4%; Score 62; DB 5; Length 322;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 211 MFLMRHGTTRT 223
    |||:|||||
OY 3 MFLSRHTGTVRT 15

RESULT 5
ID 042902 PRELIMINARY; PRT; 557 AA.
AC 042902;
DT 01-JUN-1998 (TREMREL. 06, CREATED)
DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)
DE ASPARAGINE SYNTHETASE.
GN SPB119.10.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972H-;
RA WOOD V., RAJANDREAM M.A., BARRELL B.G., SKELTON J., CHURCHER C.M.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AL022117; E1263968; -.
SQ SEQUENCE 557 AA; 63240 MW; 1805CB15 CRC32;

Query Match
Best Local Similarity 56.5%; Score 61; DB 3; Length 557;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 335 TPNYLSRK 343
    |||:|||||
OY 1 TPNYLSRN 9

RESULT 6
ID P93167 PRELIMINARY; PRT; 581 AA.
AC P93167;
DT 01-MAY-1997 (TREMREL. 03, CREATED)
DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)
DE ASPARAGINE SYNTHETASE 2 (EC 6.3.5.4) (ASPARAGINE SYNTHASE
   (GLUTAMINE-HYDROLYSING)).
DE (GLUTAMINE-HYDROLYSING)).
GN AS2.
OS GLYCINE MAX (SOYBEAN).
OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC FABALES; FABACEAE; PAPILIONOIDEAE; GLYCINE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CENTURY; TISSUE-MATURE LEAF;
RX MEDLINE: 97188563.

```

RA HUGHES C.A., BEARD H., MATTHEWS B.F.;
 PT "Molecular cloning and expression of two cDNAs encoding asparagine
 synthetase in soybean."
 RL PLANT MOL. BIOL. 33:301-311(1997).
 CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARATE + L-GLUTAMINE - AMP +
 CC DIPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
 DR EMBL: U77678; G1778370;
 DR PFAM: PF00310; GATase_2; 1.
 DR PFAM: PF00733; Asn_synthase; 1.
 DR MENDEL: 8141; Glima;1042;1.
 KW LIGASE.
 SQ SEQUENCE 581 AA; 65609 MW; 90AF9E2F CRC32;
 Query Match 56.5%; Score 61; DB 10; Length 581;
 Best Local Similarity 77.8%; Pred. No. 5.50e-01;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 DB 322 TPEFILSRK 330
 QY 1 TPEFILSRN 9
 RESULT 7
 ID 06329 PRELIMINARY; PRT; 585 AA.
 AC 06329;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE ASPARAGINE SYNTHETASE.
 GN AS.
 OS ELAEGANUS UMBELLATA.
 OC EUKARYOTA; VIRIDIPHYTA; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC RHANALES; ELAEGANACEAE; ELAEGANUS.
 RN 11
 RP SEQUENCE FROM N.A.
 RC TISSUE-ROOT NODULE;
 RA KIM H.-B., AN C.-S.;
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF061740; G3132675;
 DR MENDEL: 585 AA; 65878 MW; 8A12777D CRC32;
 SQ SEQUENCE 585 AA; 65878 MW; 8A12777D CRC32;
 Query Match 56.5%; Score 61; DB 10; Length 585;
 Best Local Similarity 77.8%; Pred. No. 5.50e-01;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 DB 322 TPEFILSRK 330
 QY 1 TPEFILSRN 9
 RESULT 8
 ID 024483 PRELIMINARY; PRT; 586 AA.
 AC 024483;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE ASPARAGINE SYNTHETASE.
 OS MEDICAGO SATIVA (ALFALFA).
 OC EUKARYOTA; VIRIDIPHYTA; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC FABALES; FABACEAE; PAPILIONOIDEAE; MEDICAGO.
 RN 11
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 97432147.
 RA SHI L., TMARY S.N., YOSHIOKA H., GREGGSON R.G., MILLER S.S.,
 RA SAMC D.A., GANTT J.S., UNKEFER P.J., VANCE C.P.;
 RT "Nitrogen assimilation in alfalfa: isolation and characterization of
 RT an asparagine synthetase gene showing enhanced expression in root
 RT nodules and dark-adapted leaves."
 RL PLANT CELL 9:1339-1356(1997).
 DR EMBL: U89923; G2522320;
 DR PFAM: PF00310; GATase_2; 1.

DR PFAM: PF00733; Asn_synthase; 1.
 SQ SEQUENCE 586 AA; 66448 MW; 63C8B6BA CRC32;
 Query Match 56.5%; Score 61; DB 10; Length 586;
 Best Local Similarity 77.8%; Pred. No. 5.50e-01;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 DB 323 TPEFILSRK 331
 QY 1 TPEFILSRN 9
 RESULT 9
 ID P93618 PRELIMINARY; PRT; 586 AA.
 AC P93618;
 DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
 DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE ASPARAGINE SYNTHETASE (EC 6.3.5.4) (ASPARAGINE SYNTHASE
 DE (GLUTAMINE-HYDROLYSING)) (ASPARAGINE SYNTHETASE
 DE (GLUTAMINE-HYDROLYSING)).
 GN VFAST.
 OS VICIA FABA (BROAD BEAN).
 OC EUKARYOTA; VIRIDIPHYTA; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC FABALES; FABACEAE; PAPILIONOIDEAE; VICIA.
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-KLEINE THUNDER; TISSUE-ROOT NODULE;
 RA KUESTER H., ALBUS U., FUEHLING M., TCHERKOVA S.A., TIKHONOVITCH I.A.,
 RA FUEHLER A., PERLICK A.M.;
 RL PLANT SCI 124:89-95(1997).
 CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARATE + L-GLUTAMINE - AMP +
 CC DIPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
 CC EMBL: 272354; E240118;
 DR PFAM: PF00310; GATase_2; 1.
 DR PFAM: PF00733; Asn_synthase; 1.
 DR MENDEL: 11022; Vicia;1042;1.
 KW LIGASE.
 SQ SEQUENCE 586 AA; 66236 MW; AD8BD927 CRC32;
 Query Match 56.5%; Score 61; DB 10; Length 586;
 Best Local Similarity 77.8%; Pred. No. 5.50e-01;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 DB 323 TPEFILSRK 331
 QY 1 TPEFILSRN 9
 RESULT 10
 ID Q40328 PRELIMINARY; PRT; 586 AA.
 AC Q40328;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE ASPARAGINE SYNTHETASE.
 OS MEDICAGO SATIVA (ALFALFA).
 OC EUKARYOTA; VIRIDIPHYTA; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC FABALES; FABACEAE; PAPILIONOIDEAE; MEDICAGO.
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. SARANAC;
 RA GANTT S.;
 RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: I40337; G1184265;
 DR PFAM: PF00310; GATase_2; 1.
 DR PFAM: PF00733; Asn_synthase; 1.
 DR MENDEL: 9036; MEDsa;1042;1.
 SQ SEQUENCE 586 AA; 66462 MW; 962C3A3F CRC32;
 Query Match 56.5%; Score 61; DB 10; Length 586;

Best Local Similarity 77.8%; Pred. No. 5,50e+01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 323 TPFLLSRK 331
|||:|:
QY 1 TPFLLSRN 9

RESULT 11
ID 058871 PRELIMINARY; PRT; 148 AA.
AC 058871;
DT 01-AUG-1998 (TREMREL. 07, CREATED)
DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)
DE 148AA LONG HYPOTHETICAL PROTEIN.
GN PH163.
OS PYROCOCUS HORIKOSHII.
UC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OT3;
RX MEDLINE; 98344137.
RA KAMABAYASHI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOTAYA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAHATA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YANAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROSE T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT "Complete Sequence and Gene Organization of the Genome of a
RT Hyper-thermophilic Archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA RES. 5:55-76(1998).
DR EMBL; AF000005; D1031206;
SQ SEQUENCE 148 AA; 17401 MW; 6B16BC02 CRC32;

Query Match 54.6%; Score 59; DB 1; Length 148;
Best Local Similarity 50.0%; Pred. No. 1.46e+00;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 13 MYLONKGVIR 24
|||:|:
QY 3 MYLONKGVIR 14

RESULT 12
ID 015039 PRELIMINARY; PRT; 820 AA.
AC 015039;
DT 01-JAN-1998 (TREMREL. 05, CREATED)
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE KIAA0327 PROTEIN.
GN KIAA0327.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 97349984.
RA NAGASE T., ISHIKAWA K., NAKAJIMA D., CHIRA M., SEKI N., MIYAJIMA N.,
RA TANAKA A., KOTANI H., NOMURA N., OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA RES. 4:141-150(1997).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AB002325; D1021623;
DR PROSITE; PS00232; CADHERIN; 5.
DR PFAM; PF000028; cadherin; 6.
KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
SQ SEQUENCE 820 AA; 89864 MW; 3B6BCA92 CRC32;

Query Match 54.6%; Score 59; DB 4; Length 820;
Best Local Similarity 61.5%; Pred. No. 1.46e+00;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 289 TPLFOLNTGTEI 301
|||:|:
QY 1 TPFLLSRNTEGEV 13

RESULT 13
ID 010288 PRELIMINARY; PRT; 593 AA.
AC 010288;
DT 01-JUL-1997 (TREMREL. 04, CREATED)
DT 01-JUL-1997 (TREMREL. 04, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE RIBONUCLEOTIDE REDUCTASE LARGE SUBUNIT.
OS ORGYIA PSEUDOTSUGATA NUCLEAR POLYEDROSIS VIRUS.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; BACULOVIRIDAE;
OC NUCLEOPOLYEDROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97271300.
RA AHRENS C.A., RUSSELL R.R., FUNK C.J., EVANS J., HARWOOD S.,
RA ROHRMANN G.F.;
RT "The sequence of the Orgyia pseudotsugata multinnucleocapsid nuclear
RT polyhedrosis virus genome.";
RL VIROLOGI 229:381-399(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA ROHRMANN G.F.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA ROHRMANN G.F.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U75930; G1911278;
DR PFAM; PF00317; ribonucleo.red; 1.
SQ SEQUENCE 593 AA; 65952 MW; 0B19B9ED CRC32;

Query Match 53.7%; Score 58; DB 14; Length 593;
Best Local Similarity 58.3%; Pred. No. 2.36e+00;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 505 TPFLLSRNTEGEV 516
|||:|:
QY 1 TPFLLSRNTEGEV 12

RESULT 14
ID 049927 PRELIMINARY; PRT; 483 AA.
AC 049927;
DT 01-JUN-1998 (TREMREL. 06, CREATED)
DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE P54 PROTEIN.
GN P54.
OS PISUM SATIVUM (GARDEN PEA).
OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC FABALES; FABACEAE; PAPILIONOIDEAE; PISUM.
RN [1]
RP SEQUENCE FROM N.A.
RA CASTILLO J., MARQUEZ J.A., FRANCO L., BALLESTAR E., RODRIGO M.I.;
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; Y11207; E321874;
SQ SEQUENCE 483 AA; 54662 MW; D4481891 CRC32;

Query Match 52.8%; Score 57; DB 10; Length 483;
Best Local Similarity 50.0%; Pred. No. 3.79e+00;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 186 TPVIVLRDENE 197
|||:|:
QY 1 TPFLLSRNTEGEV 12

```

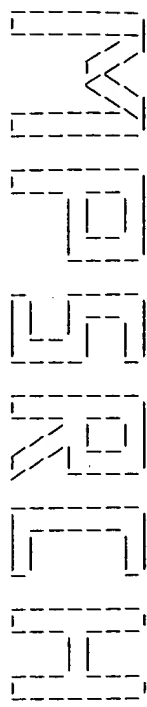
RESULT 15
ID 069782 PRELIMINARY; PRT; 300 AA.
AC 069782;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE DAPA.
GN DAPA.
OS SINORHIZOBIIUM MELILOTI.
OG PLASMID PRMGR4B.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; SINORHIZOBIIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GR4;
RA ZEKRI S., GARCIA-RODRIGUEZ F., TORO N.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AJ222715; E1286070; -.
KM PLASMID.
SQ SEQUENCE 300 AA; 32297 MM; 7992E15D CRC32;

Query Match 51.9%; Score 56; DB 2; Length 300;
Best Local Similarity 50.0%; Pred. No. 6.05e+00;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 260 AAMKLGPRGGEIR 273
QY 1 TPMFLSRNTGGEVR 14

```

Search completed: Fri Jun 11 17:31:55 1999
Job time : 127 secs.



Release 3.1A John F. Collins, Bioinformatics Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution Rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 11 17:43:07 1999; Maspar time 4.85 Seconds
65.749 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-5
Description: (1-15) from US08991628.pep
Perfect Score: 113
Sequence: 1 CECNIKVKVDVNDNP 15

Scoring table:
Gap 15
PAM 150

Searched: 170751 segs, 21266608 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

Database:

a-geneseqs5
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 18.398; Variance 59.622; scale 0.309

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	113	100.0	15	20	W04845	Self epitope of desmo	3.31e-05
2	113	100.0	999	6	R30742	Human pemphigus vulga	3.31e-05
3	102	90.3	614	19	W07908	Pemphigus vulgaris an	6.20e-04
4	89	78.8	778	21	W15489	Pemphigus foliaceus a	1.88e-02
5	72	63.7	787	17	R86865	Human protocadherin p	1.38e+00
6	57	50.4	780	24	W25634	Human cadherin-5.	4.98e+01
7	57	50.4	780	24	W25634	Human cadherin-5.	4.98e+01
8	56	49.6	833	39	W54311	Oryza sativa sequence	6.28e+01
9	56	49.6	833	39	W54311	Oryza sativa sequence	6.28e+01
10	56	49.6	1822	10	R55273	Beta subunit of integr	9.39e+01
11	54	47.8	43	11	R58865	Rat-224 cadherin part	9.39e+01
12	53	46.9	83	29	R87107	Protocadherin clone R	9.39e+01
13	53	46.9	83	29	R87107	Protocadherin clone R	9.39e+01
14	52	46.0	87	29	W55280	H. pylori ORF 06p11.1	1.25e+02
15	52	46.0	11	21	W13139	Human cadherin-5 anti	1.56e+02
16	52	46.0	132	3	R13947	E47 cDNA prod. (PE47P	1.56e+02
			162	3	R14626	Beta-lactoglobulin CO	1.56e+02

17	52	46.0	162	14	R80280	Ethyl esterified bovi	1.56e+02
18	52	46.0	178	18	R93190	Wild type beta-lactog	1.56e+02
19	52	46.0	178	18	R93191	Variant beta-lactoglo	1.56e+02
20	52	46.0	780	14	R72798	Phospholipase D.	1.56e+02
21	52	46.0	1180	34	W61273	Homo sapiens protoca	1.56e+02
22	52	46.0	3165	8	R38889	Sequence encoded by O	1.56e+02
23	51	45.1	41	11	R58875	Rat-352 cadherin part	1.96e+02
24	51	45.1	41	17	R87117	Protocadherin clone R	1.96e+02
25	51	45.1	532	24	W25640	Rat truncated cadheri	1.96e+02
26	51	45.1	532	21	W13127	Truncated rat cadheri	1.96e+02
27	51	45.1	793	21	W13132	Human cadherin-8.	1.96e+02
28	51	45.1	793	23	W25639	Full length human cad	1.96e+02
29	51	45.1	799	24	W25639	Human cadherin-8.	1.96e+02
30	51	45.1	799	24	W13126	Rat full length cadhe	1.96e+02
31	50	44.2	100	34	W38510	Streptococcus pneumon	2.44e+02
32	50	44.2	338	32	W55048	Stimuliator of iron tr	2.44e+02
33	50	44.2	658	17	R85856	WP-40 domain-contg. C	2.44e+02
34	50	44.2	2913	20	W00384	Plasmodium falciparum	2.44e+02
35	49	43.4	212	39	W73630	Human secreted protei	3.05e+02
36	49	43.4	217	1	P82052	25 kD surface antigen	3.05e+02
37	49	43.4	263	24	W13010	Segment of desmosomal	3.05e+02
38	49	43.4	304	18	R94891	CD31 fragment (domain	3.05e+02
39	49	43.4	341	31	W37907	TPV25-Pf28B fusion p	3.05e+02
40	49	43.4	360	31	W37908	TPV25-Pf28B fusion p	3.05e+02
41	49	43.4	383	18	R94892	CD31 fragment (domain	3.05e+02
42	49	43.4	440	3	R13946	E12 cDNA prod. (PE12P	3.05e+02
43	49	43.4	571	18	R94894	CD31 fragment (domain	3.05e+02
44	49	43.4	738	3	R13251	PECAM-1.	3.05e+02
45	49	43.4	842	19	W00137	Drosophila FRU 2 poly	3.05e+02

ALIGNMENTS

RESULT	1	W04845 standard; peptide; 15 AA.
ID	W04845	18-FEB-1997 (first entry)
AC	W04845	Self epitope of desmoglein 3, implicated in autoimmune disease.
DE	W04845	Tolerisation; self-epitope; antigen; autoimmune disease;
KW	W04845	autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
KW	W04845	pemphigus vulgaris; desmoglein; multiple sclerosis;
KW	W04845	herpes simplex virus; adenovirus; phosphomannomutase;
KW	W04845	human papillomavirus; Epstein-Barr virus; DNA polymerase;
KW	W04845	influenza; haemagglutinin; reovirus; sigma protein.
OS	W04845	Homo sapiens.
PN	W04845	W09627387-AL.
PD	W04845	12-SEP-1996.
PF	W04845	07-MAR-1996; U03182.
PR	W04845	07-MAR-1995; US-400796.
PA	W04845	(HARD) HARVARD COLLEGE.
PI	W04845	Strominger JL, Nuchterpennig KW;
DR	W04845	WPI; 96-425218/42.
PT	W04845	Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
PT	W04845	antigens - useful in disease treatment, and method for
PT	W04845	identification of other self and non-self antigens implicated in
PT	W04845	auto-immune disease
PS	W04845	Claim 1; Page 40; 58pp; English.
CC	W04845	Pharmaceutical preparations for tolerisation to antigens comprise
CC	W04845	either an isolated human non-collagen or non-mysin basic protein
CC	W04845	(MBP) polypeptide which is capable of tolerising an individual to an
CC	W04845	autoantigen; or an isolated human pathogen polypeptide capable of
CC	W04845	tolerising an individual to that polypeptide. In both cases, the
CC	W04845	polypeptide (whether self or non-self) includes an amino acid
CC	W04845	sequence corresponding to a sequence motif for a MHC class II
CC	W04845	protein, such as HLA-DR, which is associated with a human autoimmune
CC	W04845	disease and which binds to the polypeptide to activate autoreactive
CC	W04845	T-cells in individuals with the autoimmune disease. This peptide is
CC	W04845	derived from the human desmoglein 3 protein (amino acids 251-265)
CC	W04845	and is implicated as a self epitope in pemphigus vulgaris. Peptides
CC	W04845	derived from the human desmoglein protein are described in W0841-47.
SQ	W04845	Sequence 15 AA;
Query Match	100.0%;	Score 113; DB 20; Length 15;

Best Local Similarity 100.0%; Pred. No. 3.31e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 cecnikvdyndnf 15
OY 1 CECNIKVDVNDNFP 15

RESULT 2

ID R30742 standard; Protein: 999 AA.
AC R30742;
DT 14-JUN-1993 (first entry)
DE Human pemphigus vulgaris 130KD antigen.
KW Pemphigus vulgaris; skin disease; autoantibodies;
KM keratinocyte cell surface antigen; glycoprotein; cell adhesion.
OS Homo sapiens.
PN US798918-A.
DR 15-DEC-1992.
PF 27-NOV-1991; 798918.
PR (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Amagai M. Klaus-Kovtun V, Stanley JR;
DR N-PSDB; Q35992.
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
PS diagnostic and therapeutic uses
CC This sequence is the pemphigus vulgaris 130KD antigen. The protein
CC and its encoding DNA may be used in the diagnosis and treatment of
CC pemphigus vulgaris. It is thought that the antigen may be a cell
CC adhesion molecule.
SQ Sequence 999 AA;

Query Match 100.0%; Score 113; DB 6; Length 999;
Best Local Similarity 100.0%; Pred. No. 3.31e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 251 cecnikvdyndnf 265
OY 1 CECNIKVDVNDNFP 15

RESULT 3

ID W07908 standard; Protein: 614 AA.
AC W07908;
DT 29-JAN-1997 (first entry)
DE Pemphigus vulgaris antigen protein extracellular region.
KW Autoantibody; immunoglobulin G; IgG; fusion protein; diagnosis;
KM treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;
KW dermatology.
OS Homo sapiens.
PN J08188540-A.
PD 23-JUL-1996.
PF 30-JUN-1995; 165632.
PR 30-JUN-1994; JP-173291.
PA (NISH/) NISHIKAWA T.
DR WPI; 96-388562/39.
PT Fused protein recognised by pemphigus vulgaris auto-antibody -
PT useful to treat and diagnose pemphigus vulgaris
PS Claim 1: Page 7-9; 9pp; Japanese.
CC W07908 represents the human pemphigus vulgaris (PV) antigen
CC extracellular region. The PV antigen is produced in patients with
CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare
CC relapsing disease causing suprapasal, intra-epidermal bullae
CC (vesicles) of the skin and mucous membranes, which is fatal if
CC untreated. The PV antigen was fused to a human IgG1 hinge region
CC and the resulting fusion protein is useful to treat or diagnose
CC pemphigus vulgaris.
SQ Sequence 614 AA;

Query Match 90.3%; Score 102; DB 19; Length 614;
Best Local Similarity 93.3%; Pred. No. 6.20e-04;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 250 cecnikvdyndnf 264

OY 1 CECNIKVDVNDNFP 15

RESULT 4

ID W15489 standard; Protein: 778 AA.
AC W15489;
DT 17-JUN-1997 (first entry)
DE Pemphigus foliaceus antigen-IgG constant region fusion protein.
KW Pemphigus foliaceus; autoantibody; constant region; IgG;
KM extracellular region; antigen; hinge portion; skin;
KW dermatitis herpetiformis; fusion protein; detection; ss.
OS Chimera - Homo sapiens.
FH Key.
FT Location/Qualifiers
FT domain
FT 1..545
FT /note= "pemphigus foliaceus antigen protein"
PN J09077800-A.
PD 25-MAR-1997.
PF 12-SEP-1995; 260899.
PR 12-SEP-1995; JP-260899.
PA (NISH/) NISHIKAWA T.
DR WPI; 97-241758/22.
DR P-PSDB; T66428.
PT Pemphigus foliaceus antigen-IgG constant region fusion protein - linked
PT through the hinge region used to treat pemphigus foliaceus
PS Claim 1: Page 10-12; 17pp; Japanese.
CC This sequence represents a fused protein recognised by pemphigus
CC foliaceus patient autoantibody which comprises the constant region
CC of IgG linked to the extracellular region of pemphigus foliaceus
CC antigen protein through the hinge portion. Pemphigus foliaceus is
CC a chronic, generalised, vesicular and scaling skin eruption similar
CC to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion
CC protein is useful to treat pemphigus foliaceus. The antigen is
CC especially administered through an adjuvant upon which the fusion
CC protein is immobilised via a carrier. The fusion protein is also
CC useful for detecting pemphigus foliaceus antibodies which is useful
CC in immunodiagnosis. The fusion protein has little or no side effects.
SQ Sequence 778 AA;

Query Match 78.8%; Score 89; DB 21; Length 778;
Best Local Similarity 80.0%; Pred. No. 1.85e-02;
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 253 cecnikvdyndnf 267
OY 1 CECNIKVDVNDNFP 15

RESULT 5

ID R86865 standard; Protein: 787 AA.
AC R86865;
DT 27-AUG-1996 (first entry)
DE Human protocadherin pc3.
KW Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion;
KW catenin; therapy.
OS Homo sapiens.
PN W09600289-A1.
PD 04-JAN-1996.
PF 26-JUN-1995; U08071.
PR 27-JUN-1994; US-268161.
PA (DOHE-) DOHENY EYE INST.
PI SUZUKI S.
DR WPI; 96-068873/07.
DR N-PSDB; T03572.
PT Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
PT pc5 - involved in cell-cell adhesion and regulation activities
PS Claim 15: Page 115-119; 146pp; English.
CC R86865-R86867 represent the sequences for three protocadherins. This
CC sequence represents the human protocadherin pc3. These sequences are
CC related to cadherin, and possess cell adhesive ability. Cadherins are
CC glycosylated integral membrane proteins that are involved in cell-cell

CC adhesion. Cadherins are composed of an N-terminal extracellular domain, a C-terminal cytoplasmic domain, a membrane spanning domain, and a C-terminal cytoplasmic domain. The cytoplasmic domain interacts with the cytoskeleton through catenins and other cytoskeleton associated proteins. The cytoplasmic domain is not present in all cadherins, but in those which possess it, it is essential for the cadherin adhesive function. The cadherins which do not possess a cytoplasmic domain appear to function via a different method from those with a cytoplasmic domain. These sequences were isolated using primers 1 and 2 (see T03575 and T03576). The proteins may have regulatory functions in the cell, as well as the cell-cell adhesive properties. Antibodies produced against these sequences are useful for modulating the binding activity of these proteins, and can be used therapeutically.

Query Match 63.7%; Score 72; DB 17; Length 787;
Best Local Similarity 66.7%; Pred. No. 1.36e+00;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 331 C5VSVKVDVNDNP 345
1 CECNIKVDVNDNP 15

RESULT 6
ID W25634 standard; Protein; 780 AA.
AC W25634;
DE 03-NOV-1997 (first entry)
KW Human cadherin-5; rat; calcium-dependent cell adhesion protein;
KW Human; cadherin; rat; calcium-dependent cell adhesion protein;
KW superfamily; cytoskeleton; catenin; cancer.
OS Homo sapiens.
PI Suzuki S;
FT MISC-difference 622 location/Qualifiers
FT MISC-difference 622 /note- "encoded by AAG"

PD 08-JUL-1997. US5646250-A.
PF 17-APR-1992; 872643.
PR 19-APR-1993; US-049460.
PR 17-APR-1992; US-872643.
PR 01-NOV-1994; US-332638.
PA (DOHE-) DOHENY EYE INST.
PI Suzuki S;
DR WPI; 97-362997/33.
DR N-PSDB: T85401

PT Human and rat cadherin polypeptide(s) - mediate cell-cell adhesion
PS Claim 1; Column 69-72; 56pp; English.
CC This sequence represents human cadherin-5. The invention specifically provides details of human cadherin-5, -8, -11, -12 and -13, and rat cadherin-8, -5, -11 and -13. Cadherins are calcium-dependent cell adhesion proteins. They are glycosylated integral membrane proteins that have an N-terminal extracellular domain that determines binding specificity, a hydrophobic membrane spanning region and a C-terminal cytoplasmic domain, which is highly conserved among members of the superfamily. The C-terminal domain interacts with the cytoskeleton through catenins and other cytoskeleton-associated proteins. The novel cadherin proteins may be used in the analysis of the role of cadherins in various cancers. Sequence analysis of the cadherin proteins also allows investigation of the structure and function of cadherin. The cadherin proteins may be isolated by using anti-cadherin antibodies. These antibodies may also be used to modulate the activity of cadherin and to determine the tissue specific distribution of cadherin proteins. Each subclass of cadherins has a unique tissue distribution.

CC Pattern. 780 AA;
SQ Sequence 780 AA;

Query Match 50.4%; Score 57; DB 24; Length 780;
Best Local Similarity 81.8%; Pred. No. 4.98e+01;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 135 LKVDVNDNP 145
5 IKVKVDVNDNP 15

RESULT 7
ID W3130 standard; Protein; 780 AA.
AC W3130;
DE 13-MAY-1997 (first entry)
KW Full length human cadherin-5.
KW Ca2+ dependent; cell adhesion protein; placental; cadherin; rat;
KW brain; human; antibody; purification; determination;
KW tissue expression; binding antagonist; calcium ion.
OS Homo sapiens.
PI US5597725-A.
PD 28-JAN-1997.
PF 17-APR-1992; 872643.
PR 17-APR-1992; US-872643.
PR 19-APR-1993; US-049460.
PR 26-JAN-1994; US-188228.
PA (DOHE-) DOHENY EYE INST.
PI Suzuki S;
DR WPI; 97-108328/10.
DR N-PSDB: T61921.
PT Antibodies to cadherin proteins - useful as cadherin antagonists, etc.
PS Claim 5; Columns 75-78; 59pp; English.
CC The present sequence is full length human cadherin-5, which is a Ca2+ dependent cell adhesion protein. The human cadherin cDNA was isolated from a placental cDNA library, using probes based on homologous rat cadherin cDNA.
CC Antibodies or fragments that specifically bind the human cadherin can be used to purify the cadherin, determine its tissue expression and antagonise its ligand/antiligand binding activities.

Query Match 50.4%; Score 57; DB 21; Length 780;
Best Local Similarity 81.8%; Pred. No. 4.98e+01;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 135 LKVDVNDNP 145
5 IKVKVDVNDNP 15

RESULT 8
ID W84311 standard; Protein; 83 AA.
AC W84311;
DE 18-MAR-1999 (first entry)
KW Oryza sativa sequence used to exemplify the invention.
KW DNA sequence comparison.
OS Oryza sativa.
PN J10334104-A.
PD 18-DEC-1998.
PR 05-MAR-1998; 053102.
PR 31-MAR-1997; JP-079586.
PA (HITA) HITACHI LTD.
DR WPI; 99-110965/10.
DR Comparison of DNA base sequences - using accurate observations of insertions and deletions

PT Disclosure: Page 16-17; 24pp; Japanese.
PS The present sequence is used to demonstrate the method of the invention. The specification describes the comparison of DNA base sequences. The method involves 3 steps of: (1) division of a first and second nucleic acid sequence into base groups of three base lengths, and translation into amino acids; (2) all base and amino acid insertions and deletions of the two nucleic acid sequences are observed and the amino acid sequences encoded by these nucleic acid sequences are compared; and (3) adjacent bases and amino acids are compared and step (1) repeated after shifting the translation frame by one base towards the 3' end or CC shifting the translation frame one base towards the 3' end, but CC encompassing four bases, and not translating the second or third base. CC The new method is useful for comparing DNA base sequences.

SQ Sequence 83 AA;

Query Match 49.6%; Score 56; DB 39; Length 83;

Bst	Local similarity	40.0%;	Pred.	No. 6.28e+01;
Matches	6;	Conservative	4;	Mismatches 5; Indels 0; Gaps 0;
Dd	68 ccsilvlyideafr	82		
OY	1 CECNKKVQDVNDNP	15		
Query Match	49.6%;	Score 56;	DB 10;	Length 1822;
Best Local Similarity	63.6%;	Pred.	No. 6.28e+01;	
Matches	7;	Conservative	3;	Mismatches 1; Indels 0; Gaps
Dd	650 ecnfkymvde	660		
OY	2 ECNIKVKDVPD	12		
RESULT	10			
T	R58865 standard; protein; 43 AA.			
AC	R58865;			
DT	17-APR-1995 (first entry)			
DE	Rat-224 cadherin partial sequence.			
KW	Cadherin; cell adhesion molecule.			

PN Rattus rattus.
 PN M0941960-A.
 PD 07-JUN-1994.
 PF 23-DEC-1993; 012588.
 PR 29-DEC-1992; US-998003.
 PA (DOHE-) DOHENY EYE INST.
 PI Suzuki S;
 DR MPI: 94-293849/36.
 DR N-PSDB: 068957.
 PT Polynucleotide sequences encoding new proto:cadherins - useful
 PT for modulating natural binding and regulating activities.
 PS Example; Page 38; 114pp; English.
 CC Two regions of conserved AA sequence, one from the middle of the
 CC third cadherin extracellular subdomain (EC-3) and the other from the
 CC C-terminus of the fourth extracellular subdomain (EC-4) were
 CC identified. The corresp. degenerate oligos (068949, 068950) were
 CC designed for use as PCR primers. PCR was carried out on a rat brain
 CC cDNA prep. Two major bands of about 450 bps and 130 bps were found.
 CC The 450 bp band corresponded to the expected length between the two
 CC primer sites, but the 130 bp band could not be predicted from any
 CC of the previously identified cadherin sequences. The 450 bp and 130
 CC bp bands were extracted and sequenced. Nineteen novel partial cDNA
 CC clones were isolated. The DNA and deduced AA sequences of the
 CC clones (including sequences corresp. to the PCR primers) are given
 CC in Q68951-Q68965 and R58860-R58878. The deduced AA sequences of the
 CC cDNA clones are homologous to, but distinct from the known
 CC cadherins. The cadherins described thus far have highly conserved
 CC short AA sequences in the EC-3 including the consensus sequence
 CC D-Y-E or D-E-F located at the middle region of the subdomain and
 CC the consensus sequence in R58879 or R58880 at its end, while the
 CC corresp. sequences of other subdomains, except for the 5th extra-
 CC cellular subdomain (EC-5), are D-R-E and the sequence in R58881
 CC respectively. In contrast the deduced AA sequences of the new
 CC clones that corresp. to cadherin extracellular subdomains include
 CC the sequence D-Y-E or D-F-E at one end, but have the sequence
 CC D-X-N-D-N-X-P-X-F instead of R58879 or R58880 at the other end.
 CC The polypeptides encoded by the partial clones are homologous to
 CC previously identified cadherins but did not show significant
 CC homology to any other sequences in Genbank. Therefore, the partial
 CC cDNAs appear to comprise a new subclass of cadherin-related
 CC molecules.
 S0 Sequence 43 AA;
 Db 29 cklkxydvndnap 43
 | | | | | | | | | |
 1 CECNKKYADVNDNFP 15
 Query Match 47.8%; Score 54; DB 11; Length 43;
 Best Local Similarity 66.7%; Pred. No. 9.93e+01;
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Oy 1 CECNKKYADVNDNFP 15
 RESULT 11
 ID R87107 standard; Peptide; 43 AA.
 NC R87107;
 DE 28-AUG-1996 (first entry)
 DI Protocadherin clone RAT-224.
 KW Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion;
 KW catenin; therapy; clone.
 OS Rattus rattus.
 PN M09600289-A1.
 PD 04-JAN-1996.
 PF 26-JUN-1995; 008071.
 PR 27-JUN-1994; US-268161.
 PA (DOHE-) DOHENY EYE INST.
 PI Suzuki S;
 DR MPI: 96-068873/07.
 DR N-PSDB: 103582.
 PT Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat
 PT pc3. INVOLVED IN cell-cell adhesion and regulation activities
 PS Example 1; Page 40; 146pp; English.
 CC R87100-R87120 represent partial fragments of the rat protocadherin
 CC sequence. The cDNAs encoding these sequences were isolated after

CC screening a rat brain cDNA preparation with the primers shown in T03575
CC and T03576. The primers were constructed from portions of the amino acid
CC sequences of the third and fourth extracellular domains of published
CC cadherin sequences. The full length cDNA sequence encoding rat
CC protocadherin p05 is represented in T03574. The cytoplasmic domain of
CC cadherin interacts with the cytoskeleton through catenins and other
CC cytoskeleton associated proteins. The cytoplasmic domain is not present
CC in all cadherins, but in those which possess it, it is essential for the
CC cadherins adhesive function. The cadherins which do not possess a
CC cytoplasmic domain appear to function via a different method from those
CC with a cytoplasmic domain. These protein sequences are involved in
CC cell-cell adhesion. These sequences may have regulatory functions in the
CC cell, as well as the cell-cell adhesive properties. Antibodies produced
CC against these sequences are useful for modulating the binding activity of
CC these protocadherins, and can be used therapeutically.

Query Match 47.88; Score 54; DB 17; Length 43;
Best Local Similarity 66.78; Pred. No. 9.93e+01;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 29 ckllykvdvndnap 43
| | | | | | | | | |
QY 1 CECNIXKVDVNDNFP 15

RESULT 12

ID W55470 standard; Protein; 83 AA.

AC W55470;

DE H. pylori ORF 06ap1119_16594193_f1_9 secreted protein.

KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;

KM Identification; binding compound; bacteria; life cycle; activator;

OS Helicobacter pylori.

PN W09737044-A1.

PD 09-OCT-1997.

PF 27-MAR-1997; U05223.

PR 06-DEC-1996; US-761318.

PR 29-MAR-1996; US-625811.

PR 02-APR-1996; US-758731.

PR 25-OCT-1996; US-736905.

PR 28-OCT-1996; US-738859.

PA (ASTR) ASTRA AB.

PI Alm RA, Smith D;

PI WPI: 97-503122/46.

DR N-PSDB; V24679.

PT Helicobacter pylori nucleic acid sequences and encoded

PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori

PT infection and for diagnosis of H. pylori infection

PS Claim 14,94; Page 677-678; 1145pp; English.

CC This sequence is a H. pylori secreted protein.

CC The protein may be used in a vaccine to prevent or treat H. pylori

CC infection or to identify H. pylori polypeptide binding compounds,

CC useful as potential H. pylori life cycle activators or inhibitors. The

CC DNA and probes derived from it may be used for the identification of

CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic

CC acid sequences complementary to the DNA act as antisense sequences and

CC can be used to prevent the translation of H. pylori mRNA. Antibodies

CC against the protein can be used in immunoassays to evaluate the abundance

CC and distribution of H. pylori-specific antigens. The genomic sequence of

CC H. pylori (ATCC 55679) was determined from overlapping contigs generated

CC by mechanically shearing the bacterial DNA. The sequences were analysed

CC for ORF of at least 180 nucleotides, and the predicted coding regions

CC defined by computer evaluation. To identify likely H. pylori antigens for

CC vaccine development, the amino acid sequences predicted from various ORF

CC were analysed for significant homology to other known or exported

CC membrane proteins. Having identified and determined the sequences of

CC interest, particular regions can be isolated from H. pylori by PCR

CC amplification for recombinant polypeptide production, e.g. in E. coli

CC hosts.

CC Sequence 83 AA;

Query Match 46.98; Score 53; DB 29; Length 83;
Best Local Similarity 50.08; Pred. No. 1.25e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 43 cncnikip1 52
| | | | | | | | | |
QY 1 CECNIXKVDV 10

RESULT 13

ID W55280 standard; Protein; 87 AA.

AC W55280;

DE 02-JUL-1998 (first entry)

DE H. pylori ORF 11ge103090rf7 protein.

KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;

KM Identification; binding compound; bacteria; life cycle; activator;

KW Inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.

OS Helicobacter pylori.

PN W09737044-A1.

PD 09-OCT-1997.

PF 27-MAR-1997; U05223.

PR 06-DEC-1996; US-761318.

PR 29-MAR-1996; US-625811.

PR 02-APR-1996; US-758731.

PR 25-OCT-1996; US-736905.

PR 28-OCT-1996; US-738859.

PA (ASTR) ASTRA AB.

PI Alm RA, Smith D;

PI WPI: 97-503122/46.

DR N-PSDB; V24689.

PT Helicobacter pylori nucleic acid sequences and encoded

PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori

PT infection and for diagnosis of H. pylori infection

PS Claim 14; Page 516; 1145pp; English.

CC This sequence is a H. pylori protein of unspecified function.

CC The protein may be used in a vaccine to prevent or treat H. pylori

CC infection or to identify H. pylori polypeptide binding compounds, The

CC useful as potential H. pylori life cycle activators or inhibitors. The

CC DNA and probes derived from it may be used for the identification of

CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic

CC acid sequences complementary to the DNA act as antisense sequences and

CC can be used to prevent the translation of H. pylori mRNA. Antibodies

CC against the protein can be used in immunoassays to evaluate the abundance

CC and distribution of H. pylori-specific antigens. The genomic sequence of

CC H. pylori (ATCC 55679) was determined from overlapping contigs generated

CC by mechanically shearing the bacterial DNA. The sequences were analysed

CC for ORF of at least 180 nucleotides, and the predicted coding regions

CC defined by computer evaluation. To identify likely H. pylori antigens for

CC vaccine development, the amino acid sequences predicted from various ORF

CC were analysed for significant homology to other known or exported

CC membrane proteins. Having identified and determined the sequences of

CC interest, particular regions can be isolated from H. pylori by PCR

CC amplification for recombinant polypeptide production, e.g. in E. coli

CC hosts.

CC Sequence 87 AA;

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

RESULT 14

ID W13139 standard; Protein; 11 AA.

AC W13139;

DE 14-MAR-1997 (first entry)

DE Human cadherin-5, antigenic epitope (residues 242-252).

KW Ca2+ dependent; cell adhesion protein; cadherin; human; antibody;

KW purification; determination; epitope; tissue expression;

KW binding antagonist; calcium ion; antigen.

OS Homo sapiens.
 PN US5397725-A.
 PD 28-JAN-1997.
 PF 17-APR-1992; 872643.
 PR 17-APR-1992; US-872643.
 PR 19-APR-1993; US-049460.
 PR 26-JAN-1994; US-188228.
 PA (DOHE-) DOHERTY EYE INST.
 PI Suzuki S;
 DR MPI; 97-108328/10.
 PT Antibodies to cadherin proteins - useful as cadherin antagonists,
 etc.
 PS Claim 5: Column 112; 59pp; English.
 CC The present sequence is an antigenic epitope from human cadherin-5,
 CC which is a Ca²⁺ dependent cell adhesion protein. Antibodies or
 CC fragments that specifically bind the epitope can be used to purify
 CC the cadherin, determine its tissue expression and antagonise its
 CC ligand/antiligand binding activities.
 SQ Sequence 11 AA;

Query Match 46.0%; Score 52; DB 21; Length 11;
 Best Local Similarity 54.5%; Pred. No. 1.56e+02;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

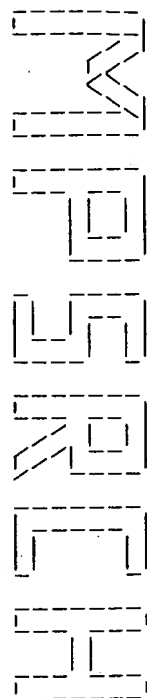
Db 1 vtqgndndfp 11
 : : 1:111111
 OY 5 IKVKDVNDNF 15

RESULT 15
 ID R13947 standard; Protein; 132 AA.
 AC R13947;
 DT 28-NOV-1991 (first entry)
 DE E47 CDNA prod. (PE47P).
 KM Immunoglobulin: enhancer-binding factor; E12; E47; E2A;
 KM chromosomal translocation; leukemia; fusion protein.
 OS Homo sapiens.
 PN W09113172-A.
 PD 05-SEP-1991.
 PF 22-FEB-1991; U01168.
 PR 23-FEB-1990; US-484063.
 PA (STRD) IRLAND STANFORD JR UNIV.
 PA (WHIT-) WHITEHEAD INST BIOMED RE.
 PI Cleary ML, Wellentin JD, Baltimore D, Murre C, Mccaw P;
 DR MPI; 91-281484/38.
 DR N-PSDB; Q13671.
 PT Detection of t(1;19) break-point-associated genes E2A and p1 -
 PT in chromosomal translocation, and prods. useful in diagnosis and
 PT therapy of human neoplasm, esp. acute lymphoblastic leukaemia
 PS Disclosure; Fig 1D; 104pp; English.
 CC The E2A gene (see Q13669), encoding immunoglobulin enhancer-binding
 CC factors E12 and E47 (see Q13670 and Q13671), is localised to the
 CC breakpoint of a consistently recurring chromosomal translocation
 CC present in many acute leukemias and is structurally altered by most
 CC t(1;19) chromosomal translocations. The translocation results in
 CC synthesis of a fusion mRNA (see Q13672 for fusion cDNA) that crosses
 CC the breakpoint between chromosome 1 and 19.
 CC The CDNs of E12 and E47 encode nearly identical proteins contg.
 CC a region that appears to represent a novel helix-loop-helix DNA-
 CC binding and dimerisation motif. The motif shares significant
 CC homology with domains in several proteins involved in the control
 CC of differentiation and proliferation in various cellular lineages.
 CC See also Q13669-75.
 SQ Sequence 132 AA;

Query Match 46.0%; Score 52; DB 3; Length 132;
 Best Local Similarity 40.0%; Pred. No. 1.56e+02;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 82 vrvrdineaf 91
 : : 1:1:1:1:1
 OY 5 IKVKDVNDNF 14

Search completed: Fri Jun 11 17:44:58 1999
 Job time : 111 secs.



Release 3.1A John F. Collins, Bioinformatics Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPERCH_p protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Jun 11 17:38:59 1999; Maspar time 4.29 Seconds
Tabular output not generated. 140.011 Million cell updates/sec

Title: >US-08-991-628-5
Description: (1-15) from US08991628.pep
Perfect Score: 113
Sequence: 1 CECNIRKVDNDNFP 15

Scoring table: PAM 150
Gap 15

Searched: 122810 segs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r60
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 25.852; Variance 36.674; scale 0.705

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	113	100.0	999	1	IJHUG3	desmoglein 3 precursor	3.28e-12
2	89	78.8	1043	1	IJBOD1	desmoglein 1 precursor	1.28e-06
3	89	78.8	1049	1	IJHUG1	desmoglein 1 precursor	1.28e-06
4	66	58.4	171	2	S52588	hypothetical protein	9.35e-02
5	63	55.8	394	2	H64448	polyferredoxin - Meth	3.57e-01
6	60	53.1	809	1	IJBODD	desmocollin 2b precu	1.31e+00
7	60	53.1	863	1	IJBODC	desmocollin 2a precu	1.31e+00
8	60	53.1	896	2	I45858	desmocollin - bovine	1.31e+00
9	58	51.3	527	2	S62484	hypothetical protein	3.06e+00
10	58	51.3	832	2	S55396	Li-cadherin - human	3.06e+00
11	58	51.3	1324	2	S51622	cut3 protein - f1aslo	4.65e+00
12	57	50.4	235	2	E70378	DNA replication prote	4.65e+00
13	57	50.4	784	1	IJHUC5	cadherin 5 precursor	4.65e+00
14	57	50.4	829	1	I46536	ksp-cadherin - rabbt	4.65e+00
15	57	50.4	847	1	IJHUBD	desmocollin 3b precu	4.65e+00
16	57	50.4	901	1	IJHUBA	desmocollin 3a precu	4.65e+00
17	57	50.4	1979	2	C71622	hypothetical protein	7.02e+00
18	56	49.6	413	1	DEACNT	D-nopaline dehydrogen	7.02e+00
19	56	49.6	787	2	S68699	potassium channel pro	7.02e+00
20	56	49.6	964	2	JC5545	integrin beta4s - hum	7.02e+00
21	56	49.6	1807	2	JC6319	integrin beta-4 chain	7.02e+00
22	56	49.6	1875	2	A36429	integrin beta-4 chain	7.02e+00
23	55	48.7	296	2	A71232	hypothetical protein	1.05e+01

24	48.7	376	2	E70361	chaperone DnaJ - Agui	1.05e+01
25	48.7	902	2	T00588	hypothetical protein	1.05e+01
26	48.7	1389	2	I50090	carboxypeptidase gp18	1.05e+01
27	47.8	249	2	G64415	hypothetical protein	1.58e+01
28	47.8	493	2	E71008	hypothetical protein	1.58e+01
29	47.8	671	2	A45730	phenol 2-monooxygenas	1.58e+01
30	47.8	1790	2	S67593	transport protein USO	1.58e+01
31	46.9	302	2	T02480	sec13-related protein	2.34e+01
32	46.9	380	2	S49116	hypothetical protein	2.34e+01
33	46.9	444	1	WHRTV	tryptophan 5-monooxyg	2.34e+01
34	46.9	646	2	E70396	histidine kinase sens	2.34e+01
35	46.9	836	2	A69550	hypothetical protein	2.34e+01
36	46.9	988	1	DJVFEP	DNA-directed DNA poly	2.34e+01
37	46.9	1986	2	S28353	probable polyketide s	3.46e+01
38	46.0	132	2	B31492	DNA-binding protein E	3.46e+01
39	46.0	135	2	D69830	hypothetical protein	3.46e+01
40	46.0	310	2	B64200	heat shock protein dn	3.46e+01
41	46.0	381	2	S48513	asparaginase (EC 3.5.	3.46e+01
42	46.0	398	2	S24802	polyferredoxin - Meth	3.46e+01
43	46.0	412	2	S46512	D-nopaline dehydrogen	3.46e+01
44	46.0	638	2	B35816	transcription regulat	3.46e+01
45	46.0	3051	2	S42373	hypothetical protein	3.46e+01

ALIGNMENTS

RESULT	1	ALIGNMENTS
ENTRY	IJHUG3	#type complete
TITLE	desmoglein 3 precursor - human	
ALTERNATE_NAMES	pemphigus vulgaris antigen	
ORGANISM	#formal_name Homo sapiens #common_name man	
DATE	30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 18-Sep-1998	
ACCESSIONS	A41088	
REFERENCE	A41088	
#authors	Amagai, M.; Klaus-Kovtun, V.; Stanley, J.R.	
#journal	Cell (1991) 67:869-877	
#title	Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a disease of cell adhesion.	
#cross-references	MIM:29269733	
#accession	A41088	
#molecule_type	mRNA	
#residues	1-999 #label AMA	
#cross-references	GB:M76482; NID:G190751; PID:G190752	
GENETICS	GDB:DSG3	
#gene	#cross-references GDB:134030; OMIM:169615	
#map_position	18q12.1-18q12.2	
CLASSIFICATION	#superfamily cadherin; cadherin repeat homology	
KEYWORDS	calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein	
FEATURE		
1-23	#domain signal sequence #status predicted #label SIG\	
24-49	#domain propeptide #status predicted #label PRO\	
50-999	#product desmoglein homolog #status predicted #label MA\	
50-615	#domain extracellular #status predicted #label EXT\	
51-157	#domain cadherin repeat homology #label CR1\	
160-267	#domain cadherin repeat homology #label CR2\	
270-383	#domain cadherin repeat homology #label CR3\	
390-495	#domain cadherin repeat homology #label CR5\	
406-508	#domain cadherin repeat homology #label CR5\	
616-639	#domain transmembrane #status predicted #label TM\	
640-999	#domain intracellular #status predicted #label INT\	
910-938	#domain desmoglein repeat #label DGL\	
937-966	#domain desmoglein repeat #label DGL2\	
110,180,545	#binding site carbohydrate (Asn) (covalent) #status predicted	
SUMMARY	#length 999 #molecular-weight 107502 #checksum 8311	
Query Match	100.0%; Score 113; Length 999;	
Best Local Similarity	100.0%; Pred. No. 3.28e-12;	
Matches	15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	

DB 251 CECNIKVDVNDNFP 265
 OY 1 CECNIKVDVNDNFP 15

RESULT 2

ENTRY 1 IJBOG1 #type complete
 TITLE desmoglein 1 precursor - bovine
 ALTERNATE_NAMES desmoglein BDGM
 ORGANISM Bos primigenius taurus #common_name cattle
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 05-Sep-1997

ACCESSIONS
 #authors S14603; A38872; A37785; S38721; A48173; S24412
 #journal S14603
 #submission Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
 #description submitted to the EMBL Data Library, March 1991
 Complete sequence of the desmoglein precursor and evidence for the existence of different desmoglein genes expressed in cell type-specific patterns.

#accession S14603
 #molecule_type mRNA
 #residues 1-1043 #label KOC
 #cross-references EMBL:X58466; NID:g306; PID:g307

REFERENCE
 #authors A38872
 Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.
 #journal Eur. J. Cell Biol. (1991) 55:200-208
 #title Complete amino acid sequence of the epidermal desmoglein precursor polypeptide and identification of a second type of desmoglein gene.

#cross-references MUID:92037656
 #accession A38872
 #molecule_type mRNA
 #residues 1-87;968-1043 #label KO2
 #cross-references GB:S64268; GB:S64270

REFERENCE
 #authors A37785
 Goodwin, L.; Hill, J.E.; Raynor, K.; Raszi, L.; Manabe, M.; Cowlin, P.
 #journal Blochem. Biophys. Res. Commun. (1990) 173:1224-1230
 #title Desmoglein shows extensive homology to the cadherin family of cell adhesion molecules.

#cross-references MUID:91097553
 #accession A37785
 #molecule_type mRNA
 #residues 44-123; 'V', 125-493 #label GOO
 #cross-references GB:M58165; NID:g162966; PID:g552318

REFERENCE
 #authors S38721
 Zimbelmann, R.
 #submission submitted to the EMBL Data Library, February 1991
 #accession S38721
 #molecule_type mRNA
 #residues 44-1043 #label ZIM
 #cross-references EMBL:X57784; NID:g436061; PID:g436062

REFERENCE
 #authors A48173
 Koch, P.J.; Walsh, M.J.; Schmelz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.
 #journal Eur. J. Cell Biol. (1990) 53:1-12
 #title Identification of desmoglein, a constitutive desmosomal glycoprotein, as a member of the cadherin family of cell adhesion molecules.

#cross-references MUID:91168965
 #accession A48173
 #molecule_type mRNA
 #residues 44-1001; 'AQPSPAR' #label KO3
 #cross-references GB:X57784
 #note this sequence has been revised in references A38872 and S38721

GENETICS
 #gene DSG1
 #classification #superfamily cadherin; cadherin repeat homology
 #keywords calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein

FEATURE
 1-23
 24-49
 50-1043
 50-548
 52-157
 160-269
 272-385
 392-491
 549-574
 575-1043
 846-875
 876-905
 906-933
 934-962
 963-1012
 110
 180,496

SUMMARY
 #length 1043 #molecular-weight 112242 #checksum 6897

Query Match
 Best Local Similarity 80.0%; Pred. No. 1,28e-06;
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 253 CECNIKVDVNDNFP 267
 OY 1 CECNIKVDVNDNFP 15

RESULT 3

ENTRY 1 IJHUG1 #type complete
 TITLE desmoglein 1 precursor - human
 ALTERNATE_NAMES desmosomal glycoprotein I
 ORGANISM Homo sapiens #common_name man
 DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 26-Feb-1998

ACCESSIONS
 #authors S16906; A39706; A61254; A61279; S16158
 #submission Buxton, R.S.
 #accession S16906
 Submitted to the EMBL Data Library, November 1990

#molecule_type mRNA
 #residues 1-1049 #label BUX
 #cross-references EMBL:X56654; NID:g30505; PID:g30506

REFERENCE
 #authors A39706
 Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Ataliotis, P.; Poynter, D.; Arneemann, J.; Rutman, A.J.; Pidsley, S.C.; Watt, F.M.; Rees, D.A.; Buxton, R.S.; Magee, A.I.
 #journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:4796-4800
 #title Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions, is related to the cadherin family of cell adhesion molecules.

#cross-references MUID:91271279
 #accession A39706
 #molecule_type mRNA
 #residues 24-1049 #label WHE
 #cross-references GB:X56654

REFERENCE
 #authors A61254
 Nilles, L.A.; Parry, D.A.D.; Powers, E.E.; Angst, B.D.; Wagner, R.M.; Green, K.J.
 #journal J. Cell Sci. (1991) 99:809-821
 #title Structural analysis and expression of human desmoglein: a cadherin-like component of the desmosome.

#accession A61254
 #molecule_type mRNA
 #residues 26-1049 #label NIL

REFERENCE
 #authors A61279
 Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arneemann, J.; Rees, D.A.; King, I.A.; Magee, A.I.
 #journal Blochem. Soc. Trans. (1991) 19:1060-1064
 #title Desmosomal glycoproteins I, II and III: novel members of the

cadherin superfamily.

#cross-references MUID:92175187

#accession A61279

#status not compared with conceptual translation

#molecule-type mRNA

#residues 1-55 #label WH3

GENETICS

#gene GDB:DSG1

#cross-references GDB:126563; OMIM:125670

#map_position 18q12.1-18q12.2

CLASSIFICATION #superfamily cadherin; cadherin repeat homology

KEYWORDS calcium binding; cell adhesion; duplication; glycoprotein;

transmembrane protein

FEATURE

1-23 #domain signal sequence #status predicted #label SIG
 24-49 #domain propeptide #status predicted #label PRO
 50-1049 #product desmoglein #status predicted #label MAT
 50-548 #domain extracellular #status predicted #label EXT
 52-157 #domain cadherin repeat homology #label CR1
 160-269 #domain cadherin repeat homology #label CR2
 272-385 #domain cadherin repeat homology #label CR3
 392-493 #domain cadherin repeat homology #label CR4
 509-530 #region serine/threonine-rich
 549-569 #domain transmembrane #status predicted #label TM
 572-1049 #domain intracellular #status predicted #label INT
 840-869 #domain desmoglein repeat #label DG1
 870-899 #domain desmoglein repeat #label DG2
 900-927 #domain desmoglein repeat #label DG3
 928-956 #domain desmoglein repeat #label DG4
 969-1019 #region glycine/serine-rich
 110,180 #binding_site carboxylate (Asn) (covalent) #status predicted

SUMMARY

#length 1049 #molecular-weight 113715 #checksum 4482

Query Match 78.8%; Score 89; DB 1; Length 1049;

Best Local Similarity 80.0%; Pred. No. 1.28e-06;

Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB

253 CECNIRKVDNDNP 267

QY 1 CECNIRKVDNDNP 15

RESULT

4

#type complete

hypothetical protein SPAC11E11.03c - fission yeast

(Schizosaccharomyces pombe)

#formal_name Schizosaccharomyces pombe

16-May-1996 #sequence_revision 13-Mar-1997 #text_change

31-Oct-1997

DATE

ACCESSIONS

REFERENCE

#authors

#submision

#accession

#status

#molecule-type DNA

#residues 1-171 #label MCL

#cross-references EMBL:Z67999; NID:q1067216; PID:q1067219

GENETICS

#map_position 1L

CLASSIFICATION #superfamily fos/jun DNA-binding domain homology

FEATURE

5-45 #domain fos/jun DNA-binding domain homology #label FJD

SUMMARY #length 171 #molecular-weight 19348 #checksum 9937

Query Match

58.4%; Score 66; DB 2; Length 171;

Best Local Similarity 35.7%; Pred. No. 9.35e-02;

Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

DB

76 COCSVKIRSVLTDF 89

QY 1 CECNIRKVDNDNP 14

RESULT

5

#type complete

polyferredoxin - Methanococcus jannaschii

#formal_name Methanococcus jannaschii

13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change

10-Oct-1997

DATE

ACCESSIONS

REFERENCE

#authors

Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
 R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton,
 R.A.; Goeysne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;
 Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
 Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glöck, A.;
 Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann,
 J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson,
 J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
 K.M.; Hurst, M.A.; Kalne, B.P.; Borodovsky, M.; Klenk,
 H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
 Science (1996) 273:1058-1073
 Complete genome sequence of the methanogenic archaeon,
 Methanococcus jannaschii.

#cross-references MUID:96337999

#accession H64448

#status

#molecule-type DNA

#residues

#cross-references GB:U67560; GB:U77117; NID:q1591813; PID:q1591821;

TIGR:M01193; PID:q1511192

CLASSIFICATION

#superfamily polyferredoxin 6x2[4Fe-4S]; ferredoxin 2[4Fe-4S]

homology

FEATURE

6-54

66-119

135-189

203-258

272-324

335-389

SUMMARY

#length 394

#molecular-weight 43342 #checksum 4356

Query Match

55.8%; Score 63; DB 2; Length 394;

Best Local Similarity 53.8%; Pred. No. 3.57e-01;

Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

DB

83 CPGRKKVDDDF 95

QY 3 CNIKRVKVDNDNP 15

RESULT

6

#type fragment

desmocollin 2b precursor - bovine (fragment)

epithelial type 2 desmocollin subform II

#formal_name Bos primigenius taurus #common name cattle

30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change

05-Sep-1997

DATE

ACCESSIONS

REFERENCE

#authors

#journal

#title

Complexity and expression patterns of the desmosomal

cadherins.

#cross-references MUID:92108053

#accession B41799

#molecule-type mRNA

#residues 1-809 #label KOC

#cross-references GB:M81190; NID:q163757; PID:q163759

#experimental_source muzzle

```

GENETICS      ##note
               #gene
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS       alternative splicing; calcium binding; cell adhesion;
               duplication; glycoprotein; transmembrane protein
FEATURE
  1-89
  90-809      #domain propeptide #status predicted #label PRO\
  90-645      #product desmocollin 2b #status predicted #label MAR\
  92-197      #domain extracellular #status predicted #label EXT\
  200-309     #domain cadherin repeat homology #label CR1\
  312-423     #domain cadherin repeat homology #label CR2\
  426-526     #domain cadherin repeat homology #label CR3\
  527-534     #domain cadherin repeat homology #label CR4\
  537-534     #domain cadherin repeat homology #label CR5\
  646-672     #domain transmembrane #status predicted #label TM\
  673-809     #domain intracellular #status predicted #label INT\
  120,346,495,579 #binding_site carbohydrate (asn) (covalent) #status
                    predicted
SUMMARY
  #length 809 #checksum 9747
Query Match      53.1%; Score 60; DB 1; Length 809;
Best Local Similarity 61.5%; Pred. No. 1,31e+00;
Matches          8; Conservative 3; Mismatches 2; Indels 0; Gaps 0
Db              295 CIIIEDVNDNLP 307
QY              3 CNIKKVDVNDNFP 15
               1 1 1 1 1 1 1 1
               1 1 1 1 1 1 1 1
RESULT          7
ENTRY          IUBODC #type fragment
TITLE          desmocollin 2a precursor - bovine (fragment)
ALTERNATE_NAMES epithelial type 2 desmocollin subform I
ORGANISM       #formal name Bos primigenius taurus #common name cattle
DATE          30-Jun-1993 #sequence revision 30-Jun-1993 #text change
               03-Sep-1997
ACCESSIONS     A41799
REFERENCE      Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Troyanovsky,
               R.; Franke, W.W.
               Proc. Natl. Acad. Sci. U.S.A. (1992) 89:353-357
               Complexity and expression patterns of the desmosomal
               cadherins.
               #cross-references MUID:92108053
               #accession A41799
               #molecule_type mRNA
               #residues 1-863 #label KOC
               #cross-references GB:M81190; NID:g163757; PID:g163758
               #experimental_source Muzzle
               #note
               #note
               264-Gln and 333-Gln were also found
GENETICS      #gene
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
KEYWORDS       alternative splicing; calcium binding; cell adhesion;
               duplication; glycoprotein; phosphoprotein; transmembrane
               protein
FEATURE
  1-89
  90-863      #domain propeptide #status predicted #label PRO\
  90-645      #product desmocollin 2a #status predicted #label MAR\
  92-197      #domain extracellular #status predicted #label EXT\
  200-309     #domain cadherin repeat homology #label CR1\
  312-423     #domain cadherin repeat homology #label CR2\
  426-526     #domain cadherin repeat homology #label CR3\
  527-534     #domain cadherin repeat homology #label CR4\
  537-534     #domain cadherin repeat homology #label CR5\
  646-672     #domain transmembrane #status predicted #label TM\
  673-863     #domain intracellular #status predicted #label INT\
  120,346,495,579 #binding_site carbohydrate (asn) (covalent) #status
                    predicted\
  #binding_site phosphate (ser) (covalent) #status

```

```

SUMMARY #length 863 #checksum 1299 predicted
Query Match 53.1%; Score 60; DB 1; Length 863;
Best Local Similarity 61.5%; Pred. No. 1,31e+00;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 295 CINIYKDVNDMP 307
1 1 1 1 1 1 1 1 1
0Y 3 CINIYKDVNDMP 15

RESULT 8
ENTRY 145858 #type complete
TITLE desmocollin - bovine
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 19-Dec-1997 #sequence_revision 19-Dec-1997 #text_change 13-Mar-1998

ACCESSIONS 145858
REFERENCE 145858
#authors Yue, K.K.; Holton, J.L.; Clarke, J.P.; Hyam, J.L.; Hashimoto, T.; Childrey, M.A.; Garrod, D.R.
#journal J. Cell Sci. (1995) 108:2163-2173
#title Characterisation of a desmocollin isoform (bovine DSC3) exclusively expressed in lower layers of stratified epithelia.
#cross_references MUID:95403557
#status 145858 preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-896 #label YUE
#cross_references GB:L33774; NID:9914820; PID:9914821
GENERICS
#gene DSC3
#introns 831/3
CLASSIFICATION #superfamily cadherin; cadherin repeat homology
SUMMARY #length 896 #molecular_weight 99687 #checksum 6602

Query Match 53.1%; Score 60; DB 2; Length 896;
Best Local Similarity 69.2%; Pred. No. 1,31e+00;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 340 CIIYKDSNDMP 352
1 1 1 1 1 1 1 1 1
0Y 3 CIIYKDSNDMP 15

RESULT 9
ENTRY 562484 #type complete
TITLE hypothetical protein SPAC4c8.07c - fission yeast
ORGANISM #formal_name Schizosaccharomyces pombe
DATE 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 31-Oct-1997

ACCESSIONS 562484
REFERENCE 562445
#authors Badcock, K.; Churcher, C.M.
#submission submitted to the EMBL Data Library, October 1995
#accession 562484
#status preliminary
#molecule_type DNA
#residues 1-527 #label BAD
#cross_references EMBL:Z56276; NID:g1022345; PID:g1022352
GENERICS
#map_position 1L
#introns 125/1; 158/2; 437/3; 520/3
SUMMARY #length 527 #molecular_weight 59613 #checksum 8195

Query Match 51.3%; Score 58; DB 2; Length 527;
Best Local Similarity 42.9%; Pred. No. 3,06e+00;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 275 ECP1ATKTINEYP 288

```

||| | :||| |
OY 2 ECNIKVKVDVNDNF 15

RESULT 10
ENTRY S55396 #type complete
TITLE LI-cadherin - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 10-Sep-1997

ACCESSIONS
REFERENCE S55396
#authors Bettinger, A.; Kreft, B.; Flegler, C.; Dlouhy, B.; Berndorff, D.; Gossens, R.; Tauber, R.
#submission submitted to the EMBL Data Library, December 1994
#description Molecular cloning of human LI-cadherin: evidence for a novel type of cadherin within the cadherin superfamily.
#accession S55396
#status preliminary
#molecule_type mRNA
#residues 1-832 #label BOE
#cross-references EMBL:X83228; NID:9854174; PID:9854175
CLASSIFICATION #superfamily cadherin repeat homology
FEATURE 455-566 #domain cadherin repeat homology #label CR3
SUMMARY #length 832 #molecular_weight 92207 #checksum 9645

Query Match 51.3%; Score 58; DB 2; Length 832;
Best Local Similarity 64.3%; Pred. NO. 3.06e+00;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 325 EIHVKVKVDVNDNF 338
OY 2 ECNIKVKVDVNDNF 15

RESULT 11
ENTRY S51622 #type complete
TITLE cut3 protein - fission yeast (Schizosaccharomyces pombe)
ORGANISM #formal_name Schizosaccharomyces pombe
DATE 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 09-Sep-1997

ACCESSIONS
REFERENCE S51622
#authors Saka, Y.; Sutan, T.; Yamashita, Y.; Saitoh, S.; Takeuchi, M.; Nakaseko, Y.; Yanagida, M.
#journal EMBO J. (1994) 13:4938-4952
#title Fission yeast cut3 and cut14, members of a ubiquitous protein family, are required for chromosome condensation and segregation in mitosis.
#cross-references MIMD:95045386
#accession S51622
#status preliminary
#molecule_type DNA
#residues 1-1324 #label SAK
SUMMARY #cross-references EMBL:D30788; NID:9577659; PID:d1007025; PID:9603501
#length 1324 #molecular_weight 150593 #checksum 3330

Query Match 51.3%; Score 58; DB 2; Length 1324;
Best Local Similarity 42.9%; Pred. NO. 3.06e+00;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 200 CDVEITREKVSDF 213
OY 1 CECNIKVKVDVNDNF 14

RESULT 12
ENTRY E70378 #type complete
TITLE DNA replication protein Dnac - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change 08-May-1998

ACCESSIONS
REFERENCE E70378
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
#cross-references MIMD:98196666
#accession E70378
#status preliminary; nucleic acid sequence not shown; translation not shown

GENETICS
#molecule_type DNA
#residues 1-235 #label AOF
#cross-references GB:AE000713; NID:92983424; PID:92983431; GB:AE000657
#experimental_source strain VF5

SUMMARY
#gene dnac
#length 235 #molecular_weight 26934 #checksum 1332

Query Match 50.4%; Score 57; DB 2; Length 235;
Best Local Similarity 63.6%; Pred. NO. 4.65e+00;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 26 CECRFRKRDVN 36
OY 1 CECNIKVKVDVN 11

RESULT 13
ENTRY IHHUC5 #type complete
TITLE cadherin 5 precursor - human
ALTERNATE_NAMES 7B4 antigen; cadherin, endothelial-specific; VE-cadherin
ORGANISM #formal_name Homo sapiens #common_name man
DATE 30-Jun-1993 #sequence_revision 13-Sep-1996 #text_change 05-Sep-1997

ACCESSIONS
REFERENCE S49893
#authors Brevierio, F.; Cavada, L.; Corada, M.; Martin-Padura, I.; Golay, J.; Introna, M.; Lampugnani, M.G.; Dejana, E.
#submission submitted to the EMBL Data Library, June 1994
#description Molecular and functional properties of VE-cadherin (7B4/cadherin-5) a novel endothelial specific cadherin.
#accession S49893
#status preliminary
#cross-references EMBL:X79981; NID:9599833; PID:9599834
#molecule_type mRNA
#residues 1-784 #label BRE
REFERENCE S24305
#authors Suzuki, S.; Sano, K.; Tanikawa, H.
#journal Cell Regul. (1991) 2:261-270
#title Diversity of the cadherin family: evidence for eight new cadherins in nervous tissue.
#cross-references MIMD:91283540
#accession S24305
#molecule_type mRNA
#residues 5-516, 'T', 518-784 #label SUZ
REFERENCE A43418
#authors Lampugnani, M.G.; Resnati, M.; Rafteri, M.; Pigott, R.; Pisacane, A.; Houen, G.; Roco, L.P.; Dejana, E.
#journal J. Cell Biol. (1992) 118:1511-1522
#title A novel endothelial-specific membrane protein is a marker of cell-cell contacts.
#cross-references MIMD:92394977
#accession A43418
#molecule_type protein
#residues 48-60, 'X', 62, 'X', 64, 108-116, 'X', 118-123; 237-238, 'X', 240, 'X', 242-252, 'X', 254-256; 263-273, 'X', 277-283, 'X', 285; 348-355; 'X', 425-429 #label IAM
#experimental_source cultured endothelial cells
#note sequence extracted from NCBI backbone (NCBIP.113040,

NCBIP:113045, NCBIP:113047, NCBIP:113049,
NCBIP:113051, NCBIP:113054)
COMMENT Cadherins mediate calcium-dependent intercellular adhesion, and are
thought to be involved in the sorting of different cell types
during morphogenesis.

GENETICS

#gene GDB:CDH5
#cross-references GDB:134230; OMIM:601120
#map_position 16q22.1-16q22.1
#CLASSIFICATION #superfamily cadherin; cadherin repeat homology
#calcium binding; cell adhesion; duplication; glycoprotein;
transmembrane protein

FEATURE

1-25
26-47 #domain signal sequence #status predicted #label SIG\
48-784 #domain propeptide #status predicted #label PRO\
48-593 #product cadherin 5 #status predicted #label MAR\
50-151 #domain extracellular #status predicted #label EXT\
154-258 #domain cadherin repeat homology #label CR1\
261-372 #domain cadherin repeat homology #label CR2\
375-479 #domain cadherin repeat homology #label CR3\
481-587 #domain cadherin repeat homology #label CR4\
594-620 #domain cadherin repeat homology #label CR5\
621-784 #domain transmembrane #status predicted #label TMN\
736-753 #domain intracellular #status predicted #label INT\
61,112,157,362,442, #binding_site carboxylate (Asn) (covalent) #status
523,535 predicted

SUMMARY

#length 784 #molecular-weight 87516 #checksum 7459
Query Match 50.4%; Score 57; DB 1; Length 784;
Best Local Similarity 81.8%; Pred. No. 4.65e+00;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 139 IKYDVNDNMP 149
||| |||||
OY 5 IKYKVDNDNMP 15

RESULT 14
ENTRY 146536 #type complete
TITLE Ksp-cadherin - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change
14-Feb-1997

ACCESSIONS 146536
REFERENCE 146536
#authors Thomson, R.B.; Igarashi, P.; Biemesderfer, D.; Kim, R.;
Abu-Alfa, A.; Soleimani, M.; Atkinson, P.S.
#journal J. Biol. Chem. (1995) 270:17594-17601
#title Isolation and cDNA cloning of Ksp-cadherin, a novel
kidney-specific member of the cadherin multigene family.
#cross-references MIM:95340560
#accession I46536
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-829 #label THO
#cross-references EMBL:U28945; NID:9902885; PID:9902886
SUMMARY #length 829 #molecular-weight 88827 #checksum 5058

Query Match 50.4%; Score 57; DB 2; Length 829;
Best Local Similarity 53.3%; Pred. No. 4.65e+00;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 433 CEVAVTVTDVNDNMP 447
||:| |||||:
OY 1 CECNKKVDVNDNMP 15

RESULT 15
ENTRY ITHUBB #type complete
TITLE desmocollin 3b precursor - human

ALTERNATE NAMES desmosomal glycoprotein III
ORGANISM #formal_name Homo sapiens #common_name man
DATE 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change
12-Jun-1998

ACCESSIONS A40390, S16464
REFERENCE A40390
#authors Parker, A.E.; Wheeler, G.N.; Arnemann, J.; Pidsley, S.C.;
Ataliotis, P.; Thomas, C.L.; Rees, D.A.; Magee, A.I.;
Buxton, R.S.
#journal J. Biol. Chem. (1991) 266:10438-10445
#title Desmosomal glycoproteins II and III. Cadherin-like junctional
molecules generated by alternative splicing.

#cross-references MIM:91244819
#accession A40390
#molecule_type mRNA
#residues 1-847 #label PAR
#cross-references GB:X56807
#note It is uncertain whether Met-1 is the initiator or
whether translation is initiated upstream to the
sequenced region

REFERENCE A43032
#authors Buxton, R.S.; Cowlin, P.; Franke, W.W.; Garrod, D.R.; Green,
K.J.; King, I.A.; Koch, P.J.; Magee, A.I.; Rees, D.A.;
Stanley, J.R.; Steinberg, M.S.
#journal J. Cell Biol. (1993) 121:481-483
#title Nomenclature of the desmosomal cadherins.
#contents annotation; nomenclature

GENETICS

#gene GDB:DSC3; DSC2; DSC1; DS
#cross-references GDB:126552; OMIM:600271
#map_position 16q12.1-18q12.1
#CLASSIFICATION #superfamily cadherin; cadherin repeat homology
#alternative splicing; calcium binding; cell adhesion;
duplication; glycoprotein; phosphoprotein; transmembrane
protein

FEATURE 1-28
29-135 #domain signal sequence #status predicted #label SIG\
136-847 #domain propeptide #status predicted #label PRO\
136-695 #product desmocollin 3b #status predicted #label MAR\
138-243 #domain extracellular #status predicted #label EXT\
246-355 #domain cadherin repeat homology #label CR1\
358-471 #domain cadherin repeat homology #label CR2\
474-577 #domain cadherin repeat homology #label CR3\
578-680 #domain cadherin repeat homology #label CR4\
719-847 #domain transmembrane #status predicted #label TMN\
160,392,546,629 #domain intracellular #status predicted #label INT\
#binding_site carboxylate (Asn) (covalent) #status
predicted

SUMMARY #length 847 #molecular-weight 93768 #checksum 3131

Query Match 50.4%; Score 57; DB 1; Length 847;
Best Local Similarity 53.8%; Pred. No. 4.65e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 341 CIINIDVDNDNMP 353
||:| |||||:
OY 3 CNIKVDVNDNMP 15

Search completed: Fri Jun 11 17:40:26 1999
Job time : 87 secs.

KW CELL ADHESION: SIGNAL: TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
 KW CALCIUM-BINDING; REPEAT.
 FT SIGNAL 1 23 POTENTIAL.
 FT PROPEP 24 49 POTENTIAL.
 FT CHAIN 50 999 DESMOGLEIN 3.
 FT DOMAIN 50 615 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 616 640 POTENTIAL.
 FT DOMAIN 641 999 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 159 158 CADHERIN 1.
 FT REPEAT 159 268 CADHERIN 2.
 FT REPEAT 269 383 CADHERIN 3.
 FT REPEAT 386 499 CADHERIN 4.
 FT REPEAT 910 935 DESMOGLEIN REPEAT 1.
 FT REPEAT 936 966 DESMOGLEIN REPEAT 2.
 FT CARBOHYD 110 110 POTENTIAL.
 FT CARBOHYD 180 180 POTENTIAL.
 FT CARBOHYD 459 459 POTENTIAL.
 FT CARBOHYD 545 545 POTENTIAL.
 SQ SEQUENCE 999 AA; 107503 MM; 48918FAE CRC32;

Query Match 100.0%; Score 113; DB 1; Length 999;
 Best Local Similarity 100.0%; Pred. No. 6,78e-15;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 251 CECNIRKVDNDNFP 265
 OY 1 CECNIRKVDNDNFP 15

RESULT 2
 ID DSG1-BOVIN STANDARD; PRT; 1043 AA.
 AC 003763;
 DT 01-OCT-1993 (REL. 27, CREATED)
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN I) (DGI).
 GN DSG1.
 OS BOS TAURUS (BOVINE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-MUZZLE EPITHELIUM.
 RA KOCH P.J., GOLDSCHMIDT M.D., ZIMBELMANN R., FRANK W.W.;
 RL SUBMITTED (MAR-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE OF 44-1043 FROM N.A.
 RC TISSUE-MUZZLE EPITHELIUM.
 RA KOCH P.J., WALSH M.J., SCHMELZ M., GOLDSCHMIDT M.D.,
 RA ZIMBELMANN R., FRANK W.W.;
 RT Identification of desmoglein, a constitutive desmosomal
 glycoprotein, as a member of the cadherin family of cell adhesion
 molecules.
 RT EUR. J. CELL BIOL. 53:1-12(1990).
 RN [3]
 RP REVISIONS, AND SEQUENCE OF 101-123.
 RA MEDLINE: 92037656.
 RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R.,
 RA FRANK W.W.;
 KI Complete amino acid sequence of the epidermal desmoglein precursor
 RT polypeptide and identification of a second type of desmoglein gene.
 RT EUR. J. CELL BIOL. 55:200-208(1991).
 RN [4]
 RP SEQUENCE OF 44-493 FROM N.A.
 RA MEDLINE: 91097553.
 RA GOODWIN L., HILL J.E., RAYNOR K., RASZI L., MANBE M., COMIN P.;
 RT Desmoglein shows extensive homology to the cadherin family of cell
 adhesion molecules.
 RT BIOCHEM. BIOPHYS. RES. COMMUN. 173:1224-1230(1990).
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
 CC INVOLVED IN THE INTERACTION OF PLASMA PROTEINS AND INTERMEDIATE
 CC FILAMENTS MEDIATING CELL-CELL ADHESION.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN
 CC -1- TISSUE SPECIFICITY: EPIDERMIS, MUZZLE, TONGUE AND OESOPHAGUS.
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE
 CC DESMOSOMAL SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----

DR EMBL: X58466; G307; -
 DR EMBL: X57784; G436062; -
 DR EMBL: M58165; G552318; -
 DR PIR: S14603; IUBOGL.
 DR PROSITE: PS00232; CADHERIN; 2.
 DR PFAM: PF00028; cadherin; 3.
 DR HSSP: P09803; 1EDH.
 DR CELL ADHESION: SIGNAL: TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
 KW CALCIUM-BINDING; REPEAT.

FT SIGNAL 1 23 POTENTIAL.
 FT PROPEP 24 49 POTENTIAL.
 FT CHAIN 50 1043 DESMOGLEIN 1.
 FT DOMAIN 50 548 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 549 573 POTENTIAL.
 FT DOMAIN 574 1043 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 50 158 CADHERIN 1.
 FT REPEAT 159 270 CADHERIN 2.
 FT REPEAT 271 385 CADHERIN 3.
 FT REPEAT 386 498 CADHERIN 4.
 FT REPEAT 499 845 DESMOGLEIN REPEAT 1.
 FT REPEAT 846 875 DESMOGLEIN REPEAT 2.
 FT REPEAT 876 905 DESMOGLEIN REPEAT 3.
 FT REPEAT 906 933 DESMOGLEIN REPEAT 4.
 FT DOMAIN 934 1012 DESMOGLEIN REPEAT 5.
 FT CARBOHYD 110 110 GLY/SER-RICH.
 FT CARBOHYD 180 180 POTENTIAL.
 FT CARBOHYD 496 496 POTENTIAL.
 FT CONFLICT 124 124 I -> V (IN REF. 4).
 SQ SEQUENCE 1043 AA; 112243 MM; 13898584 CRC32;

Query Match 78.8%; Score 89; DB 1; Length 1043;
 Best Local Similarity 80.0%; Pred. No. 2.56e-08;
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

DB 253 CECNIRKVDNDNFP 267
 OY 1 CECNIRKVDNDNFP 15

RESULT 3
 ID DSG1-HUMAN STANDARD; PRT; 1049 AA.
 AC 002413;
 DT 01-OCT-1993 (REL. 27, CREATED)
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN I) (DGI).
 GN DSG1.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KERATINOCYTES;
 RX MEDLINE: 91271279.
 RA WHEELER G.N., PARKER A.E., THOMAS C.L., ATALIONIS P., POYNTER D.,
 RA ARNEBARN J., RUTMAN A.J., FIDSLIEY S.C., WATT F.M., REES D.A.,
 RA BOXTON R.S., WAGEE A.I.;

RT "Desmosomal glycoprotein DGI, a component of intercellular desmosome
RT junctions, is related to the cadherin family of cell adhesion
RT molecules.";
RL PROC. NATL. ACAD. SCI. U.S.A. 88:4796-4800(1991).
CC -1- FUNCTION: COMPONENT OF INTRACELLULAR DESMOsome JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL AND OESOPHAGUS.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE
CC DESMOsome SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X56654; G30506; -
DR PIR: S16906; IJHUG1.
DR MIM: 125670; -
DR PROSITE: PS00232; CADHERIN; 2.
DR PFAM: PF00028; cadherin; 4.
DR HSSP: P09803; 1EDH.
KW CELL ADHESION; SIGNAL; TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
KW CALCIUM-BINDING; REPEAT.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 49 POTENTIAL.
FT CHAIN 50 1049 DESMOGLEIN 1.
FT DOMAIN 50 545 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 546 570 POTENTIAL.
FT DOMAIN 571 1049 CYTOPLASMIC (POTENTIAL).
FT REPEAT 50 158 CADHERIN 1.
FT REPEAT 159 270 CADHERIN 2.
FT REPEAT 271 385 CADHERIN 3.
FT REPEAT 386 497 CADHERIN 4.
FT REPEAT 813 839 DESMOGLEIN REPEAT 1.
FT REPEAT 840 869 DESMOGLEIN REPEAT 2.
FT REPEAT 870 899 DESMOGLEIN REPEAT 3.
FT REPEAT 900 927 DESMOGLEIN REPEAT 4.
FT REPEAT 928 956 DESMOGLEIN REPEAT 5.
FT DOMAIN 969 1019 GLY/SER-RICH.
FT CARBOHYD 36 36 POTENTIAL.
FT CARBOHYD 110 110 POTENTIAL.
FT CARBOHYD 180 180 POTENTIAL.
SQ SEQUENCE 1049 AA; 113715 MW; FDD79961 CRC32;
Query Match 78.8%; Score 89; DB 1; Length 1049;
Best Local Similarity 80.0%; Pred. No. 2,56e-08;
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

RP SEQUENCE FROM N.A.
RX MEDLINE; 96140438.
RA WATANABE Y., YAMAMOTO M.;
RT "Schizosaccharomyces pombe pcr1+ encodes a CREB/ATF protein involved
RT in regulation of gene expression for sexual development.";
RL MOL. CELL. BIOL. 16:704-711(1996).
RN [2]
RN SEQUENCE FROM N.A.
RA KON N., KRAWCHUK M.D., WARREN B.G., SMITH G.R., WAHLS W.P.;
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN-972;
RA MCELAN J., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RN CHARACTERIZATION.
RX MEDLINE; 95047325.
RA WAHLS W.P., SMITH G.R.;
RT "A heteromeric protein that binds to a meiotic homologous
RT recombination hot spot: correlation of binding and hot spot
RT activity.";
RL GENES DEV. 8:1693-1702(1994).
CC -1- FUNCTION: INVOLVED IN REGULATION OF GENE EXPRESSION FOR SEXUAL
CC DEVELOPMENT. BINDS AND ACTIVATES MEIOTIC RECOMBINATION HOT SPOT
CC ADB6-W26.
CC -1- SUBUNIT: HETERODIMER OF PCRI/MTS2 AND ATF1/MTS1.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D63667; G119496; -
DR EMBL: U87870; G1839252; -
DR EMBL: Z67999; E1168681; -
DR PROSITE: PS00036; BZIP_BASIC; 1.
DR PFAM: PF00170; bzip; 1.
DR HSSP: P05412; 1FOS.
KW TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING; NUCLEAR PROTEIN;
KW MEIOSIS.
FT DNA_BIND 12 32 BASIC MOTIF.
FT DOMAIN 42 66 LEUCINE-ZIPPER.
SQ SEQUENCE 171 AA; 19348 MW; 0D025155 CRC32;
Query Match 58.4%; Score 66; DB 1; Length 171;
Best Local Similarity 35.7%; Pred. No. 1,22e-02;
Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

DB 253 CECNIRILDVNDNIP 267
OY 1 CECNIRKVDVNDNF 15
RESULT 4
ID PCRI_STANDARD; PRT; 171 AA.
AC 009926;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PCRI OR MTS2 OR SPAC21E1.03C.
GN SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIZASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]

DB 76 CQCSVKIRSLTDF 89
OY 1 CECNIRKVDVNDNF 14
RESULT 5
ID DSC2_BOVIN_STANDARD; PRT; 863 AA.
AC P33545;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE DESMOCOLLIN 2A/2B PRECURSOR (EPITHELIAL TYPE 2 DESMOCOLLIN)
DE (FRAGMENT).
GN DSC2.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE-MOZZLE EPITHELIUM;
 RX MEDLINE; 92108053.
 RA KOCH P.J., GOLDSCHMIDT M.D., ZIMBELMANN R., TROYANOVSKY R.,
 FRANK W.W.,
 RT "Complexity and expression patterns of the desmosomal cadherins,"
 RL "PROC. NATL. ACADE. SCI. U.S.A. 89:353-357(1992)."
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
 CC IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
 CC MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
 CC POSITONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
 CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: OESOPHAGUS AND RUMEN. WEAKLY IN EPITHELIA AND
 CC CARDIAC MUSCLE.
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -1- ALTERNATIVE PRODUCTS: FORMS 2A (SHOWN HERE) AND 2B ARE PRODUCED BY
 CC ALTERNATIVE SPLICING OF THE SAME GENE.
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE
 CC DESMOSOMAL SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M81190; G163758; -
 CC EMBL; M81190; G163759; -
 CC PIR; A41799; IJBODC.
 CC PIR; B41799; IJBODC.
 CC DR PROSITE; PS00232; CADHERIN; 3.
 CC DR PFAM; PF00028; cadherin; 4.
 CC DR HSSP; P09803; ISOH.
 CC KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT;
 CC ALTERNATIVE SPLICING; CYTOSKELETON; CALCIUM-BINDING.
 CC FT PROPEP 1 1
 CC FT CHAIN 1 89
 CC FT DOMAIN 90 863
 CC FT TRANSMEM 90 644
 CC FT DOMAIN 645 665
 CC FT REPEAT 666 863
 CC FT REPEAT 90 197
 CC FT REPEAT 198 309
 CC FT REPEAT 310 423
 CC FT REPEAT 424 528
 CC FT REPEAT 529 644
 CC FT CARBOHYD 120 120
 CC FT CARBOHYD 346 346
 CC FT CARBOHYD 495 495
 CC FT CARBOHYD 579 579
 CC FT VARIANT 264 264
 CC FT VARIANT 333 333
 CC FT VARSPLIC 799 809
 CC FT VARSPLIC 810 863
 CC FT SEQUENCE 863 AA; 95874 MW; 2173F06E CRC32;
 CC
 CC Query Match 53.1%; Score 60; DB 1; Length 863;
 CC Best Local Similarity 61.5%; Pred. No. 2.61e-01;
 CC Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 CC
 CC Db 295 CINIENYNNLP 307
 CC QY 3 CNIKVDVNDNFP 15
 CC
 CC RESULT 6
 CC ID DSC3_BOVIN STANDARD; PRT; 896 AA.
 CC AC Q28060; Q28061; Q28176;
 CC DT 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE DESMOCOLLIN 3a/3b PRECURSOR.
 GN DSC3.
 OS BOS TAURUS (BOVINE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 NC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 95403557.
 RA YUE K.K.M., HOLTON J.L., CLARKE J.P., HYAM J.L.M., HASHIMOTO T.,
 RA CHIDGEY M.A.J., GARROD D.R.,
 RT "Characterisation of a desmocollin isoform (bovine DSC3) exclusively
 RT expressed in lower layers of stratified epithelia,"
 RL J. CELL SCI. 108:2163-2173(1995).
 RN [2]
 RP SEQUENCE OF 686-814 FROM N.A.
 RC TISSUE-EPIDERMIS;
 RX MEDLINE; 94308280.
 RA LEGAN P.K., YUE K.K.M., CHIDGEY M.A.J., HOLTON J.L., WILKINSON R.W.,
 RA GARROD D.R.,
 RT "The bovine desmocollin family: a new gene and expression patterns
 RT reflecting epithelial cell proliferation and differentiation,"
 RL J. CELL BIOL. 126:507-518(1994).
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
 CC IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
 CC MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
 CC POSITONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
 CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: STRATIFIED EPITHELIA ONLY (EPIDERMIS, TONGUE,
 CC OESOPHAGUS AND RUMEN).
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -1- ALTERNATIVE PRODUCTS: FORMS 3A AND 3B ARE PRODUCED BY ALTERNATIVE
 CC SPLICING OF THE SAME GENE.
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE
 CC DESMOSOMAL SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L33774; G914821; -
 CC EMBL; L33774; G914822; -
 CC EMBL; X75783; G433870; -
 CC DR PROSITE; PS00232; CADHERIN; 3.
 CC DR PFAM; PF00028; cadherin; 5.
 CC DR HSSP; P09803; ISOH.
 CC KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;
 CC ALTERNATIVE SPLICING; CYTOSKELETON; CALCIUM-BINDING.
 CC FT SIGNAL 1 26
 CC FT PROPEP 27 134
 CC FT CHAIN 135 896
 CC FT DOMAIN 135 690
 CC FT TRANSMEM 691 711
 CC FT DOMAIN 712 896
 CC FT REPEAT 135 242
 CC FT REPEAT 243 354
 CC FT REPEAT 355 471
 CC FT REPEAT 472 579
 CC FT DOMAIN 580 690
 CC FT VARSPLIC 832 839
 CC FT VARSPLIC 840 896
 CC FT VARSPLIC 165 165
 CC FT CARBOHYD 391 391
 CC FT CARBOHYD 546 546
 CC FT CARBOHYD 629 629
 CC FT CONFLICT 686 687
 CC
 CC Query Match 53.1%; Score 60; DB 1; Length 863;
 CC Best Local Similarity 61.5%; Pred. No. 2.61e-01;
 CC Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 CC
 CC Db 295 CINIENYNNLP 307
 CC QY 3 CNIKVDVNDNFP 15
 CC
 CC RESULT 6
 CC ID DSC3_BOVIN STANDARD; PRT; 896 AA.
 CC AC Q28060; Q28061; Q28176;
 CC DT 01-NOV-1997 (REL. 35, CREATED)
 CC VI -> EF (IN REF. 2).

DR EMBL: M68963; G325513; -
 KW DNA REPLICATION; DNA-BINDING; ATP-BINDING.
 FT NP_BIND 52 59 ATP (POTENTIAL).
 SQ SEQUENCE 780 AA; 89716 MW; 5f096841 CRC32;
 Query Match 52.2%; Score 59; DB 1; Length 780;
 Best Local Similarity 53.3%; Pred. No. 4.27e-01;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

DB 424 CENSIRKVDVNGFP 438
 1 CECNIKXVDNDNFP 15

RESULT 10
 ID DSC2_MOUSE STANDARD; PRT; 902 AA.
 AC P55292; 064734;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE DESMOCOLLIN 2A/2B PRECURSOR (EPITHELIAL TYPE 2 DESMOCOLLIN).
 GN DSC2 OR DSC3.
 OS MUS MUSCULUS (MOUSE).
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 CC RODENTIA; SCIROGNATHI; MURIDAE; MURINAE; MUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57BL/6; TISSUE-EMBRYO;
 RX MEDLINE; 95227276.
 RA LORIMER J.E., HALL L.S., CLARKE J.P., COLLINS J.E., FLEMING T.P.,
 RA GARROD D.R.,
 RT "Cloning, sequence analysis and expression pattern of mouse
 RT desmocollin 2 (DSC2), a cadherin-like adhesion molecule."
 RT MOL. MEMBRANE BIOL. 11:229-236(1994).
 RN [2]
 RP SEQUENCE OF 344-637 FROM N.A.
 RC STRAIN-C57BL/6; TISSUE-EMBRYO;
 RX MEDLINE; 95048328.
 RA BUXTON R.S., WHEELER G.N., PIDSLLEY S.C., MARSDEN M.D., ADAMS M.J.,
 RA JENKINS N.A., GILBERT D.J., COPELAND N.G.,
 RT "Mouse desmocollin (Dsc3) and desmoglein (Dsg1) genes are closely
 RT linked in the proximal region of chromosome 18."
 RT GENOMICS 21:510-516(1994).
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSE JUNCTIONS. INVOLVED
 CC IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
 CC MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
 CC POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
 CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: IN ALL EPITHELIA TESTED AND HEART.
 CC -1- DOMAIN: CALCULON MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -1- ALTERNATIVE PRODUCTS: FORMS 2A AND 2B ARE PRODUCED BY ALTERNATIVE
 CC SPLICING OF THE SAME GENE.
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE
 CC DESMOSEOMAL SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L33779; G1017465; -
 DR EMBL: L33779; G498019; -
 DR EMBL: X73865; G439671; -
 DR MGD: MGI:103221; DSC2.
 DR PROSITE: PS00232; CADHERIN; 3.
 DR PFM: PFM0028; cadherin; 4.
 DR HSP: P09803; IEDH.
 KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;

KW ALTERNATIVE SPLICING; CYTOSKELETON; CALCIUM-BINDING.
 FT SIGNAL 1 27
 FT PROPER 28 135
 FT CHAIN 136 902
 FT DOMAIN 136 694
 FT TRANSMEM 695 715
 FT DOMAIN 716 902
 FT REPEAT 136 243
 FT REPEAT 244 355
 FT REPEAT 356 471
 FT REPEAT 472 579
 FT REPEAT 580 694
 FT CARBOHYD 166 166
 FT CARBOHYD 392 392
 FT CARBOHYD 546 546
 FT CARBOHYD 629 629
 FT VARSPLIC 838 848
 FT VARSPLIC 849 902
 SQ SEQUENCE 902 AA; 99961 MW; 37FE753B CRC32;
 Query Match 52.2%; Score 59; DB 1; Length 902;
 Best Local Similarity 61.5%; Pred. No. 4.27e-01;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

DB 341 CIIIEDVDNDLP 353
 3 CNIKXVDNDNFP 15

RESULT 11
 ID YAD7_SCHPO STANDARD; PRT; 527 AA.
 AC Q09833;
 DT 01-FEB-1996 (REL. 33, CREATED)
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 59.6 KD PROTEIN C468.07C IN CHROMOSOME I.
 GN SPAC468.07C.
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
 CC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
 CC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
 CC SCHIZOSACCHAROMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RA BACOCK K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.,
 RL SUBMITTED (OCT-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- SIMILARITY: TO YEAST ENDO-EXONUCLEASE NUCL (RNC1).
 CC -1- SIMILARITY: SOME, TO THE RNA METHYLTRANSFERASE TRMA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z56276; G1022352; -
 DR PROSITE: PS01230; TRMA.1; 1.
 DR PROSITE: PS01231; TRMA.2; 1.
 KW HYPOTHETICAL PROTEIN; HYDROLASE; NUCLEASE; TRANSFERASE;
 KW METHYLTRANSFERASE.
 FT ACT_SITE 479 479
 FT ACT_SITE 479 479
 SQ SEQUENCE 527 AA; 59613 MW; 5852B845 CRC32;
 Query Match 51.3%; Score 58; DB 1; Length 527;
 Best Local Similarity 42.9%; Pred. No. 6.94e-01;
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

DB 275 ECPINKTINEELP 288
 2 CENIKXVDNDNFP 15

RESULT 12
ID CUT3_SCHPO STANDARD; PRT; 1324 AA.
AC P41004;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE CHROMOSOME SEGREGATION PROTEIN CUT3.
GN CUT3.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95045386.
RA SAKA Y., SUTANI T., YAMASHITA Y., SAITOH S., TAKEUCHI M.,
NAKASEKO Y., YAMAGIDA M.,
FISSION yeast cut3 and cut4, members of a ubiquitous protein
family, are required for chromosome condensation and segregation in
mitosis.
RT EMBL J.13:4938-4952(1994).
CC -1- FUNCTION: REQUIRED FOR CHROMOSOME CONDENSATION AND SEGREGATION
IN MITOSIS. LOCATION: NUCLEAR.
CC -1- SUBCELLULAR LOCATION: BELONGS TO THE SMC FAMILY.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
DR EMBL; D30788; G603501; -
KW MITOSIS; ATP-BINDING; COILED COIL; NUCLEAR PROTEIN.
FT NP_BIND 155 162 ATP (POTENTIAL).
FT DOMAIN 310 337 COILED COIL (POTENTIAL).
FT DOMAIN 370 628 COILED COIL (POTENTIAL).
FT DOMAIN 825 1077 COILED COIL (POTENTIAL).
FT DOMAIN 1297 1324 COILED COIL (POTENTIAL).
SQ SEQUENCE 1324 AA; 150594 MW; 76FB343D CRC32;
Query Match 51.3%; Score 58; DB 1; Length 1324;
Best Local Similarity 42.9%; Pred. No. 6,94e-01;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
Db 200 CVELTFKEVNSDF 213
1: 1 1: 1 1
QY 1 CECNIKVDNDF 14

RESULT 13
ID CAD5_MOUSE STANDARD; PRT; 783 AA.
AC P53284;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE VASCULAR ENDOTHELIAL-CADHERIN PRECURSOR (VE-CADHERIN) (CADHERIN-5).
GN CDH5.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MORINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-BRAIN CAPILLARY;
RX MEDLINE; 96141083.
RA BREIER G., BREVIARIO F., CAVEDA L., BERTHIER R., SCHNEIDER H.,
GOTSCH U., VESTWEBER D., RISAU W., DEJANA E.,
"Molecular cloning and expression of murine vascular endothelial-
cadherin in early stage development of cardiovascular system.",
BLOOD 87:630-641(1996).

CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS: CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES. THIS CADHERIN MAY PLAY A
CC IMPORTANT ROLE IN ENDOTHELIAL CELL BIOLOGY THROUGH CONTROL OF THE
CC COHESION AND ORGANIZATION OF THE INTERCELLULAR JUNCTIONS. NERVOUS
CC SYSTEM.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL; X83930; E135091; -
DR MGD; MGI:105057; CDH5.
DR PROSITE; PS00232; CADHERIN; 3.
DR PFAM; PF00028; cadherin; 5.
DR PFAM; PF01049; Cadherin_C-term; 1.
DR HSSP; P09803; LEDH.
KW CELL ADHESION; GLYCOPROTEIN; PHOSPHORYLATION; TRANSMEMBRANE;
KW CALCIUM-BINDING; REPEAT; SIGNAL.
FT SIGNAL 1 24
FT PROPEP 25 45 POTENTIAL.
FT CHAIN 46 783 VASCULAR ENDOTHELIAL-CADHERIN.
FT DOMAIN 46 592 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 593 619 POTENTIAL.
FT DOMAIN 620 783 CYTOPLASMIC (POTENTIAL).
FT REPEAT 46 148 CADHERIN 1.
FT REPEAT 149 255 CADHERIN 2.
FT REPEAT 256 370 CADHERIN 3.
FT REPEAT 371 475 CADHERIN 4.
FT REPEAT 476 592 CADHERIN 5.
FT DOMAIN 737 752 SER-RICH.
FT CARBOHYD 59 59 POTENTIAL.
FT CARBOHYD 154 154 POTENTIAL.
FT CARBOHYD 440 440 POTENTIAL.
FT CARBOHYD 522 522 POTENTIAL.
FT CARBOHYD 534 534 POTENTIAL.
SQ SEQUENCE 783 AA; 87847 MW; D0B71215 CRC32;
Query Match 50.4%; Score 57; DB 1; Length 783;
Best Local Similarity 63.6%; Pred. No. 1.12e+00;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 243 IRLDINDNFP 253
1: 1 1: 1 1 1 1 1
QY 5 IKVKVDNDF 15

RESULT 14
ID CAD5_HUMAN STANDARD; PRT; 784 AA.
AC P33151;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE VASCULAR ENDOTHELIAL-CADHERIN PRECURSOR (VE-CADHERIN) (CADHERIN-5)
(7B4 ANTIGEN) (CD144 ANTIGEN).
GN CDH5.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-ENDOTHELIAL CELLS;
RX MEDLINE; 95353875.
RA BREVIARIO F., CAVEDA L., CORADA M., MARTIN-PADURA I., NAVARRO P.,
GOLAY J., INTRONA M., GULINO D., LAMPIGNANI M.G., DEJANA E.,
"Functional properties of human vascular endothelial cadherin

RT (7B4/cadherin-5), an endothelium-specific cadherin.";
 RL ARTERIOSCLER. THROMB. VASC. BIOL. 15:1228-1239(1995).
 RN [2]
 RC SEQUENCE FROM N.A.
 RC TISSUE-PLACENTA;
 RC ALI J., MULLER W.;
 RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE OF 5-784 FROM N.A.
 RC TISSUE-BRAIN;
 RC MEDLINE: 91283540.
 RA SUZUKI S., SANO K., TANIHARA H.;
 RT "Diversity of the cadherin family: evidence for eight new cadherins
 in nervous tissue."
 RL CELL REGUL. 2:261-270(1991).
 RN [4]
 RP PARTIAL SEQUENCE.
 RC TISSUE-ENDOTHELIAL CELLS;
 RC MEDLINE: 92394977.
 RA LAMPUGNANI M.G., RESNATI M., RAITERI M., PIGOTT R., PISACANE A.,
 RA HOUEN G., RUO L.P., DELANA E.;
 RT "A novel endothelial-specific membrane protein is a marker of
 cell-cell contacts."
 RL CELL BIOL. 118:1511-1522(1992).
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
 CC MANNER IN CONNECTING CELLS. CADHERINS MAY THUS CONTRIBUTE TO THE
 CC SORTING OF HETEROGENEOUS CELL TYPES. THIS CADHERIN MAY PLAY A
 CC IMPORTANT ROLE IN ENDOTHELIAL CELL BIOLOGY THROUGH CONTROL OF THE
 CC COHESION AND ORGANIZATION OF THE INTERCELLULAR JUNCTIONS. IT
 CC ASSOCIATES WITH ALPHA-CATENIN FORMING A LINK TO THE CYTOSKELETON.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. FOUND AT CELL-CELL
 CC BOUNDARIES AND PROBABLY AT CELL-MATRIX BOUNDARIES.
 CC -1- TISSUE SPECIFICITY: ENDOTHELIAL TISSUES AND BRAIN.
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
 CC -1- DATABASE: NAME-PROV: NOTE-CD guide CD144 entry;
 CC WWW: "http://www.ncbi.nlm.nih.gov/ncbi/cd/cd144.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X79981; G598934;
 DR EMBL: U84722; G1814076;
 DR EMBL: X59796; G29593;
 DR PIR: S24305; IJHUC5.
 DR MIM: 601120;
 DR PROSITE: PS00232; CADHERIN; 3.
 DR PFM: PF00028; cadherin_5.
 DR PFM: PF01049; cadherin_C-term; 1.
 DR HSSP: P09803; 1EDH.
 KW CELL ADHESION; GLYCOPROTEIN; PHOSPHORYLATION; TRANSMEMBRANE;
 KW CALCIUM-BINDING; REPEAT; SIGNAL.
 FT SIGNAL 1 25
 FT PROPEP 26 47
 FT CHAIN 48 784
 FT DOMAIN 48 593
 FT TRANSMEM 594 620
 FT DOMAIN 621 784
 FT REPEAT 48 151
 FT REPEAT 152 258
 FT REPEAT 259 372
 FT REPEAT 373 477
 FT REPEAT 478 593
 FT DOMAIN 736 753
 FT CARBOHYD 61 61
 FT CARBOHYD 112 112
 FT CARBOHYD 157 157
 FT CARBOHYD 362 362
 FT POTENTIAL.
 FT POTENTIAL.
 FT VASCULAR ENDOTHELIAL-CADHERIN.
 FT EXTRACELLULAR (POTENTIAL).
 FT POTENTIAL.
 FT CYTOPLASMIC (POTENTIAL).
 FT CADHERIN 1.
 FT CADHERIN 2.
 FT CADHERIN 3.
 FT CADHERIN 4.
 FT CADHERIN 5.
 FT SER-RICH.
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.


FT CARBOHYD 442 442 POTENTIAL.
 FT CARBOHYD 523 523 POTENTIAL.
 FT CARBOHYD 535 535 POTENTIAL.
 FT CONFLICT 517 517 I -> T (IN REF. 1).
 SQ SEQUENCE 784 AA; 87528 MW; C2C5CD71 CRC32;
 Query Match 50.4%; Score 57; DB 1; Length 784;
 Best Local Similarity 81.8%; Pred. No. 1,128+00;
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Db 139 IKVHDVNDNP 149
 Oy 5 IKVKNVNDNP 15
 RESULT 15
 ID DSC2_HUMAN STANDARD; PRT; 901 AA.
 AC 002487;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE DESMOCOLLIN 2A/2B PRECURSOR (DESMOSOMAL GLYCOPROTEIN II AND III)
 DE (DESMOCOLLIN-3).
 GN DSC2 OR DSC3.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KERATINOCYTES;
 RC MEDLINE: 91244819.
 RA PARKER A.E., WHEELER G.N., ARNEMANN J., PIDSLEY S.C., ATALLOTIS P.,
 RA THOMAS C.L., REES D.A., MAGEE A.I., BUXTON R.S.;
 RT "Desmosomal glycoproteins II and III. Cadherin-like junctional
 RT molecules generated by alternative splicing."
 RL J. BIOL. CHEM. 266:10438-10445(1991).
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
 CC IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
 CC MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
 CC POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
 CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN EPITHELIA, MYOCARDIUM AND LYMPH
 CC NODES.
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
 CC (POTENTIAL).
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS; 2A OR DGI1 (SHOWN HERE) AND 2B OR
 CC DGI11; ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE
 CC DESMOSOMAL SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X56807; G30508; ALT_INIT.
 DR EMBL: X56807; G30509; ALT_INIT.
 DR PIR: B40380; IJHODA.
 DR PIR: A40390; IJHODA.
 DR MIM: 125645;
 DR PROSITE: PS00232; CADHERIN; 3.
 DR PFM: PF00028; cadherin; 4.
 DR HSSP: P09803; 1EDH.
 KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;
 KW ALTERNATIVE SPLICING; CYTOSKELETON; CALCIUM-BINDING; PHOSPHORYLATION.
 FT SIGNAL 1 27
 FT PROPEP 28 135
 FT CHAIN 136 901
 FT DOMAIN 136 694
 FT POTENTIAL.
 FT POTENTIAL.
 FT DESMOCOLLIN 2A/2B.
 FT EXTRACELLULAR (POTENTIAL).

```
FT TRANSMEM 695 715 POTENTIAL.
FT DOMAIN 716 901 CYTOPLASMIC (POTENTIAL).
FT REPEAT 136 243 CADHERIN 1.
FT REPEAT 244 355 CADHERIN 2.
FT REPEAT 356 471 CADHERIN 3.
FT REPEAT 472 579 CADHERIN 4.
FT REPEAT 580 694 CADHERIN 5.
FT REPEAT 34 34 POTENTIAL.
FT CARBOHYD 166 166 POTENTIAL.
FT CARBOHYD 392 392 POTENTIAL.
FT CARBOHYD 546 546 POTENTIAL.
FT CARBOHYD 629 629 POTENTIAL.
FT MOD_RES 864 864 PHOSPHORYLATION (POTENTIAL).
FT VARSPLIC 837 847 KYLCNODENH -> ESIRGHTLIRN (IN FORM
  2B/DGIII).
FT VARSPLIC 848 901 MISSING (IN FORM 2B/DGIII).
SQ SEQUENCE 901 AA: 99961 MW: CB688FC4 CRC32;
```

```
Query Match 50.48; Score 57; DB 1; Length 901;
Best Local Similarity 53.88; Pred. No. 1.12e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
```

```
DB 341 CLINIDVNDHLP 353
OY 3 CNIKVDVNDNEP 15
```

Search completed: Fri Jun 11 17:36:19 1999
Job time : 8 secs.


 ZHU
 (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

Mpsrch_pp  protein - protein database search, using Smith-Waterman algorithm
Run on:      Fri Jun 11 17:36:36 1999;  MasPar time 6.12 Seconds
Tabular output not generated.          133.859 Million cell updates/sec

```

```
Title: >US-08-991-628-5
Description: (1-15) from US08991628.pep
RefSeq Score: 113
Sequence: 1 CECNIKVKDVNDNFP 15
```

Scoring table: PAM 150
Gap 15

```
Searched: 179066 seqs, 54579741 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
```

```
Database:
  spiremb19
    1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
    5:sp_invertebrate 6:sp_mammal 7:sp_mmc 8:sp_oranelle
    9:sp_phase 10:sp_plant 11:sp_rudent 12:sp_unclassified
    13:sp_vertebrate 14:sp_virus
```

Statistics: Mean 25.706; Variance 33.200; scale 0.774

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description	Pred. No
	1	107	94.7	993	11	035902	DESMOGELIN 3 (FRAGMENT	4,84e-11
	2	65	57.5	223	3	013737	HYPOTHETICAL 25.1 KD P	6.34e-07
	3	65	57.5	519	13	090425	VENTRAL NEURAL CADHERI	6.34e-07
	4	63	55.8	384	1	058593	METHYLOLOGEN-REDICIN	1.68e-07
	5	63	55.8	2606	5	021920	RIIAB.7 PROTEIN.	1.68e-07
	6	62	54.9	379	2	008356	DNAJ PROTEIN.	2.72e-07
	7	62	54.9	501	5	008213	CO2D4.2 PROTEIN.	2.72e-07
	8	60	53.1	466	5	076322	SYNAPSIN S-SYN-SHORT (7.03e-07
	9	60	53.1	503	5	076323	SYNAPSIN S-SYN-LONG (F	7.03e-07
	10	60	53.1	814	6	077704	DESMOCOLLIN TYPE 2 (FR	7.03e-07
	11	59	52.2	659	14	069468	REPLICATION ORIGIN-BIN	1.12e+00
	12	59	52.2	2163	5	001912	SIMILARITY TO MULTIPLE	1.12e+00
	13	58	51.3	339	7	P79570	MHC CLASS I PRECURSOR.	1.78e+00
	14	58	51.3	832	4	012864	INTESTINAL PEPTIDE-ASS	1.78e+00
	15	58	51.3	832	4	015336	LT-CADHERIN.	1.78e+00
	16	58	51.3	5157	3	001135	PEPTIDE SYNTHETASE.	1.78e+00
	17	57	50.4	235	2	067056	DNA REPLICATION PROTEI	2.81e+00
	18	57	50.4	784	11	035542	CADHERIN-5.	2.81e+00
	19	57	50.4	829	6	028634	KIDNEY-SPECIFIC CADHER	2.81e+00
	20	57	50.4	915	5	062328	R0SH10.6 PROTEIN.	2.81e+00

21	56	49.6	802	10	Q388980	POTASSIUM CHANNEL (ART	4.42e+00
22	56	49.6	964	4	0146980	INTEGRIN VARIANT BETA4	4.42e+00
23	56	49.6	1752	4	015341	BETA4-INTEGRIN.	4.42e+00
24	56	49.6	1805	4	015330	BETA4-INTEGRIN.	4.42e+00
25	56	49.6	1822	4	015340	BETA4-INTEGRIN.	4.42e+00
26	55	48.7	261	3	013908	HYPOHETICAL 27.9 KD P	6.90e+00
27	55	48.7	296	1	057855	296AA LONG HYPOHETICA	6.90e+00
28	55	48.7	376	2	066921	CHAPERONE DNAJ	6.90e+00
29	55	48.7	807	1	075288	KIA0345-LIKE 13.	6.90e+00
30	55	48.7	830	11	088338	KSP-CADHERIN.	6.90e+00
31	55	48.7	902	10	064736	TZEL1.16 PROTEIN.	6.90e+00
32	55	48.7	906	5	017261	CADHERIN HOMOLOG.	6.90e+00
33	55	48.7	1387	13	057512	CARBOXYPEPTIDASE D.	6.90e+00
34	55	48.7	1389	13	030240	CARBOXYPEPTIDASE GP180	6.90e+00
35	55	48.7	2809	5	061120	G-CADHERIN.	6.90e+00
36	54	47.8	292	5	016616	B0281.8 PROTEIN.	1.07e+01
37	54	47.8	493	1	059088	493AA LONG HYPOHETICA	1.07e+01
38	54	47.8	807	10	024538	POTASSIUM CHANNEL (FRA	1.07e+01
39	54	47.8	1435	5	064997	K12C11.4 PROTEIN.	1.07e+01
40	54	47.8	1790	3	007380	HYPOHETICAL 206.5 KD	1.07e+01
41	53	46.9	83	6	002829	TPP-TYRPHAN HYDROXY	1.65e+01
42	53	46.9	350	2	P70897	VARIABLE MAJOR PROTEIN	1.65e+01
43	53	46.9	380	8	036232	MITOCHONRIAL APP9 GEN	1.65e+01
44	53	46.9	496	14	092339	POLYPROTEIN (FRAGMENT)	1.65e+01
45	53	46.9	2479	11	Q63002	MANNOSE 6-PHOSPHATE/IN	1.65e+01

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	993 AA.
AC	035902;			
DT	01-JAN-1998 (TREMBLREL. 05, CREATED)			
DT	01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	DESMOLEIN 3 (FRAGMENT).			
GN	DG3.			
OS	MUS MUSCULUS (MOUSE).			
OC	EUAROTIA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA			
OC	SCUROGNATHI; MURIDAE; MORINAE; MUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BALB/C;			
RA	ISHIKAWA H., LI K., UITTO J.;			
RL	SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.			
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).			
DR	EMBL; U86016; G2290200; -.			
DR	PROSITE; PS00232; CADHERIN; 2.			
DR	PFAM; PF00028; cadherin; 4.			
KW	CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.			
FT	NON_TER 993			
SQ	SEQUENCE 993 AA; 107888 MW; 881794BD CRC32;			
Query March	94.7%; Score 107; DB 11; Length 993;			
Best Local Similarity	86.7%; Pred. No. 4, 84e-11;			
Matches 13; Conservative	2; Mismatches 0; Indels 0; Gaps			
Db	251 CECISIKIDVNDNFP 265			
OY	1 CECNIKIKVDVNDNF 15			
RESULT	2			
ID	013737	PRELIMINARY;	PRT;	223 AA.
AC	013737;			
DT	01-JUN-1998 (TREMBLREL. 06, CREATED)			
DT	01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)			
DT	01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)			
DE	HYPOTHETICAL 25.1 KD PROTEIN C16E8.02 IN CHROMOSOME I.			
GN	SPAC16E8.02.			
OS	SCHIZOSACCHAROMYCES POMBE (YEAST).			
OC	EUAROTIA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;			
OC	SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;			

CC SCHIZOSACCHAROMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RA OLIVER K., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
 RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- SIMILARITY LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: YEAST YGL010W AND SOME, TO N.CRASSA ATP-6.
 DR EMBL: 298529; E334015;
 KM HYPOTHETICAL PROTEIN: TRANSMEMBRANE.
 FT TRANSMEM 28 48 POTENTIAL.
 FT DOMAIN 31 34 POLY-LEU.
 FT TRANSMEM 66 86 POTENTIAL.
 FT TRANSMEM 106 126 POTENTIAL.
 FT TRANSMEM 136 156 POTENTIAL.
 SQ SEQUENCE 223 AA; 25122 MW; CB8F8E6B CRC32;
 Query Match 57.5%; Score 65; DB 3; Length 223;
 Best Local Similarity 66.7%; Pred. No. 6.34e-02;
 Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 Db 168 NIKKDVGENP 179
 QY 4 NIKKDVNDNP 15
 RESULT 3 PRELIMINARY; PRT; 519 AA.
 ID 090425;
 AC 090425;
 DT 01-NOV-1996 (TREMBLER, 01, CREATED)
 DT 01-NOV-1996 (TREMBLER, 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLER, 08, LAST ANNOTATION UPDATE)
 DE VENTRAL NEURAL CADHERIN (FRAGMENT).
 OS BRACHIDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).
 CC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: ACTINOPTERYGII: NEOPTERYGII:
 CC TELEOSTEI: EUTELEOSTEI: OSTARIOPHYSI: CYPRINIFORMES: CYPRINOIDEA;
 CC CYPRINIDE: RASBORINAE; DANIO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA FRANKLIN J.L., SARGENT T.D.;
 RL DEV. DYN. 206:0-0(0).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: U04119; G1345125;
 DR PROSITE; PS00232; CADHERIN; 1.
 DR PFAM; PF00028; cadherin; 3.
 DR PFAM; PF01049; Cadherin_C-term; 1.
 KM CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
 FT NON-TER 1
 FT SEQUENCE 519 AA; 57807 MW; E6DA0079 CRC32;
 Query Match 57.5%; Score 65; DB 13; Length 519;
 Best Local Similarity 91.7%; Pred. No. 6.34e-02;
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 202 NIKKDVNDNP 213
 QY 4 NIKKDVNDNP 15
 RESULT 4 PRELIMINARY; PRT; 394 AA.
 ID 058593;
 AC 058593;
 DT 01-AUG-1998 (TREMBLER, 07, CREATED)
 DT 01-AUG-1998 (TREMBLER, 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLER, 08, LAST ANNOTATION UPDATE)
 DE METHYLOXIGEN-REDUCING HYDROGENASE POLYFERREDOXIN PROTEIN.
 CN VNUB OR KIL13.
 OS METHANOCOCCUS JANNASCHII.
 CC ARCHAEA: EURYARCHAEOTA: METHANOCOCCALES; METHANOCOCCACEAE;
 CC METHANOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DSM 2661;

RX MEDLINE; 96337999.
 RA BOLT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOGANE J.D.,
 RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
 RA OVERBEER R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLONER A.,
 RA SCOTT J.L., GEORGHEN N.S.M., WEIDMAN J.E., FUHRMANN J.L., NGUYEN D.,
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KANE B.P., BORODOVSKY M.,
 RA KLENK H.-P., FRASER C.M., SMITH H.O., MOESE C.R., VENTER J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii";
 RL SCIENCE 273:1058-1073(1996)
 CC -1- FUNCTION: FERREDOXINS ARE IRON-SULFUR PROTEINS THAT TRANSFER
 CC ELECTRONS IN A WIDE VARIETY OF METABOLIC REACTIONS.
 CC -1- THIS PROTEIN CONTAINS SEVERAL REPEATED FERREDOXIN-LIKE DOMAINS.
 CC -1- COFACTOR: BINDS TEN 4FE-4S CLUSTER.
 CC -1- SUBUNIT: VNUB CONSISTS OF THE VNUA, VNUG, VNUH SUBUNITS AND A
 CC FERREDOXIN PROTEIN.
 DR EMBL: U67560; G1591821;
 DR PROSITE; PS00198; 4FE4S_FERREDOXIN; 10.
 DR PFAM; PF00037; fer4; 6.
 KM ELECTRON TRANSPORT; IRON-SULFUR; REPEAT.
 FT METAL 13 13
 FT METAL 16 16 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
 FT METAL 19 19 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
 FT METAL 23 23 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
 FT METAL 36 36 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 39 39 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 42 42 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 46 46 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 73 73 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
 FT METAL 76 76 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
 FT METAL 79 79 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
 FT METAL 83 83 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
 FT METAL 101 101 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
 FT METAL 104 104 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
 FT METAL 107 107 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
 FT METAL 111 111 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
 FT METAL 142 142 IRON-SULFUR 5 (4FE-4S) (BY SIMILARITY).
 FT METAL 145 145 IRON-SULFUR 5 (4FE-4S) (BY SIMILARITY).
 FT METAL 148 148 IRON-SULFUR 5 (4FE-4S) (BY SIMILARITY).
 FT METAL 152 152 IRON-SULFUR 5 (4FE-4S) (BY SIMILARITY).
 FT METAL 171 171 IRON-SULFUR 6 (4FE-4S) (BY SIMILARITY).
 FT METAL 174 174 IRON-SULFUR 6 (4FE-4S) (BY SIMILARITY).
 FT METAL 177 177 IRON-SULFUR 6 (4FE-4S) (BY SIMILARITY).
 FT METAL 181 181 IRON-SULFUR 6 (4FE-4S) (BY SIMILARITY).
 FT METAL 210 210 IRON-SULFUR 7 (4FE-4S) (BY SIMILARITY).
 FT METAL 213 213 IRON-SULFUR 7 (4FE-4S) (BY SIMILARITY).
 FT METAL 216 216 IRON-SULFUR 7 (4FE-4S) (BY SIMILARITY).
 FT METAL 220 220 IRON-SULFUR 7 (4FE-4S) (BY SIMILARITY).
 FT METAL 240 240 IRON-SULFUR 8 (4FE-4S) (BY SIMILARITY).
 FT METAL 243 243 IRON-SULFUR 8 (4FE-4S) (BY SIMILARITY).
 FT METAL 246 246 IRON-SULFUR 8 (4FE-4S) (BY SIMILARITY).
 FT METAL 250 250 IRON-SULFUR 8 (4FE-4S) (BY SIMILARITY).
 FT METAL 306 306 IRON-SULFUR 9 (4FE-4S) (BY SIMILARITY).
 FT METAL 309 309 IRON-SULFUR 9 (4FE-4S) (BY SIMILARITY).
 FT METAL 312 312 IRON-SULFUR 9 (4FE-4S) (BY SIMILARITY).
 FT METAL 316 316 IRON-SULFUR 9 (4FE-4S) (BY SIMILARITY).
 FT METAL 343 343 IRON-SULFUR 10 (4FE-4S) (BY SIMILARITY).
 FT METAL 346 346 IRON-SULFUR 10 (4FE-4S) (BY SIMILARITY).
 FT METAL 349 349 IRON-SULFUR 10 (4FE-4S) (BY SIMILARITY).
 FT METAL 353 353 IRON-SULFUR 10 (4FE-4S) (BY SIMILARITY).
 SQ SEQUENCE 394 AA; 43342 MW; 9C45BC4F CRC32;
 Query Match 55.8%; Score 63; DB 1; Length 394;
 Best Local Similarity 53.8%; Pred. No. 1.66e-01;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 Db 83 CPPIKRVDDNP 95
 QY 3 CNIKKDVNDNP 15

```

RESULT 5
ID 021920 PRELIMINARY; PRT; 2606 AA.
AC 021920; Q21927;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE R11A8.7 PROTEIN.
UN R11A8.7
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA BARBIL S.;
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIRKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA CUMMINGS P.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; 270686; E1348745; JOINED.
DR EMBL; 270310; E1348745; JOINED.
DR EMBL; 270686; E1348615; JOINED.
SQ SEQUENCE 2606 AA; 285415 MW; F4D767A8 CRC32;

Query Match 55.8%; Score 63; DB 5; Length 2606;
Best Local Similarity 53.8%; Pred. No. 1.68e-01;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

DB 965 CAISVRDMESNFP 977
OY 3 CNIKVKVDNDF 15

```

```

DR PROSITE; PS00637; DNAJ_CXXCXGXG; 1.
DR PFAM; PF00226; DnaJ_1.
DR PFAM; PF00684; DnaJ_CXXCXGXG; 1.
DR CHAPERONE; DNA REPLICATION.
SQ SEQUENCE 379 AA; 40992 MW; 5D2BC6EE CRC32;

Query Match 54.9%; Score 62; DB 2; Length 379;
Best Local Similarity 38.5%; Pred. No. 2.72e-01;
Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

DB 44 QCEIKFKEINEAY 56
OY 2 EGNIKVKVDNDF 14

RESULT 7
ID 002213 PRELIMINARY; PRT; 501 AA.
AC 002213;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE C02D4.2 PROTEIN.
GN C02D4.2
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA BAYNES C.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIRKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
RN [3]
RP SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC EMBL; 281031; E1343679; JOINED.
DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
DR G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN.
SQ SEQUENCE 501 AA; 56404 MW; 18F1C57F CRC32;

Query Match 54.9%; Score 62; DB 5; Length 501;
Best Local Similarity 54.5%; Pred. No. 2.72e-01;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

DB 336 CKYKVKDVKED 346
OY 3 CNIKVKVDNDF 13

RESULT 8
ID 076322 PRELIMINARY; PRT; 466 AA.
AC 076322;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SYNAPSIN S-SYN-SHORT (FRAGMENT).
OS LOLLIGO PALATIS (LOLLIGO SQUID).
OC EUKARYOTA; METAZOA; MOLUSCA; CEPHALOPODA; COLEOIDEA; TEUTHOIDA;
OC MYOPSIDA; LOLLIGINIDAE; LOLLIGO.
RN [1]
RP SEQUENCE FROM N.A.

```

RA HILFIER S., SCHWEIZER F.E., KAO H.-T., CZERNIK A.J., GREENGARD P.,
 RA AUGUSTINE G.J.,
 RT "Two Sites of Action for Synapsin Domain E in Regulating
 RT Neurotransmitter Release."
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF055373; G3273724; -
 LT NON_TER 1
 SO SEQUENCE 466 AA; 52133 MW; C751B337 CRC32;
 Query Match 53.1%; Score 60; DB 5; Length 466;
 Best Local Similarity 35.7%; Pred. No. 7.03e-01;
 Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
 Db 287 CKCDHVKIGNY 300
 QY 1 CECNKKVDNDNF 14
 RESULT 9
 ID 076323; PRELIMINARY; PRT; 503 AA.
 AC 076323;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE SYNAPSTIN S-SYN-LONG (FRAGMENT).
 OS LOULIGO PEALETTI (LONGFIN SQUID).
 OC EUKARYOTA; METAZOA; MOLUSCA; CEPHALOPODA; COLEOIDEA; TEUTHOIDA;
 OC MYOPSIDA; LOLIGINIDAE; LOLIGO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA HILFIER S., SCHWEIZER F.E., KAO H.-T., CZERNIK A.J., GREENGARD P.,
 RA AUGUSTINE G.J.,
 RT "Two Sites of Action for Synapsin Domain E in Regulating
 RT Neurotransmitter Release."
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF055374; G3273726; -
 LT NON_TER 1
 SO SEQUENCE 503 AA; 55738 MW; 693971D9 CRC32;
 Query Match 53.1%; Score 60; DB 5; Length 503;
 Best Local Similarity 35.7%; Pred. No. 7.03e-01;
 Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
 Db 287 CKCDHVKIGNY 300
 QY 1 CECNKKVDNDNF 14
 RESULT 10
 ID 077704; PRELIMINARY; PRT; 814 AA.
 AC 077704;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE DESMOCOLLIN TYPE 2 (FRAGMENT).
 OS DSC2.
 GN CANIS FAMILIARIS (DOG).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC CARNIVORA; FISSIPEDIA; CANIDAE; CANIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE: 98380123.
 RA ROBERTS G.A., BURDETT I.D., PIDSLEY S.C., KING I.A., MAGEE A.I.,
 RA BUXTON R.S.,
 RT "Antisense expression of a desmocollin gene in MDCK cells alters
 RT desmosome plaque assembly but does not affect desmoglein
 RT expression."
 RL EUR. J. CELL. BIOL. 76:192-203(1998).
 CC -1. SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: AF002299; E1315148; -
 DR PROSITE: P500332; CADHERIN; 3.
 KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
 LT NON_TER 1
 SO SEQUENCE 814 AA; 91081 MW; BE0007A8 CRC32;

FT CHAIN 101 814 DESMOCOLLIN TYPE 2.
 SQ SEQUENCE 814 AA; 91081 MW; BE0007A8 CRC32;
 Query Match 53.1%; Score 60; DB 6; Length 814;
 Best Local Similarity 61.5%; Pred. No. 7.03e-01;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 Db 306 CINIIDDVNDNF 318
 QY 3 CNIKKVDNDNF 15
 RESULT 11
 ID 069468; PRELIMINARY; PRT; 659 AA.
 AC 069468;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE REPLICATION ORIGIN-BINDING PROTEIN (FRAGMENT).
 GN HDREO.
 OS HUMAN HERPESVIRUS TYPE 6.
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE; BETAHERPESVIRINAE;
 OC ROSLOVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-01102;
 RA MEDLINE: 95027704.
 RA NICHOLAS J.,
 RT "Nucleotide sequence analysis of a 21-kbp region of the genome of
 RT human herpesvirus-6 containing homologues of human cytomegalovirus
 RT major immediate-early and replication genes."
 RL VIROLOGY 204:738-750(1994).
 DR EMBL: U13194; G662096; -
 LT NON_TER 1
 SO SEQUENCE 659 AA; 75779 MW; 65BEBAB0 CRC32;
 Query Match 52.2%; Score 59; DB 14; Length 659;
 Best Local Similarity 53.3%; Pred. No. 1.12e+00;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 Db 303 CENSTKVDVNGFP 317
 QY 1 CECNKKVDNDNF 15
 RESULT 12
 ID 001912; PRELIMINARY; PRT; 2163 AA.
 AC 001912;
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE SIMILARITY TO MULTIPLE CADHERIN-TYPE REPEATS.
 GN R10R2.1.
 OS CAENORABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTERA; RHABDITIA; RHABDITIDA;
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORABDITIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA MEDLINE: 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
 RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCGURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULLIVAN J.,
 RA THERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHIDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans".
 RT NATURE 368:32-38(1994).

RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA DU 2., GATUNG S.;
 RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: AF003388; G2088852; -;
 DR PROSITE: PS00232; CADHERIN; 8.
 DR PFAM: PF00028; cadherin; 14.
 KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
 SQ SEQUENCE 2163 AA; 238609 MW; 6C85C652 CRC32;

Query Match 52.2%; Score 59; DB 5; Length 2163;
 Best Local Similarity 53.3%; Pred. No. 1.12e+00;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 1629 CKCHVIVLDNDNP 1643

QY 1 CECNIKVDVNDNF 15

RESULT 13
 ID P79570 PRELIMINARY; PRT; 339 AA.
 AC P79570;
 DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
 DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
 DT 01-JUN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE MHC CLASS I PRECURSOR.
 GN ONGO 92H.
 OS ONCORYNHUS GORBUSCHA (PINK SALMON) (HUMPAK SALMON).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
 OC TELEOSTEI; EUTELEOSTEI; PROTACANTHOPTERYGII; SALMONTIFORMES; SALMONIDAE;
 OC ONCORYNHUS.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY;
 RX MEDLINE: 97113756.
 RA KTAGIRI T., HIRONO I., AOKI T., SAKAI M.;
 KI "Isolation of major histocompatibility complex class I cDNA from pink
 RT salmon (Oncorhynchus gorbuscha).";
 RL DEV. COMP. IMMUNOL. 20:217-228(1996).
 DR EMBL: D58386; D1010196; -;
 DR PRAM: PF00047; 1q; 1.
 DR PFAM: PF00129; MHC_I; 1.
 KW SIGNAL; MHC.
 FT SIGNAL 1 19
 SQ SEQUENCE 339 AA; 38268 MW; B5224C56 CRC32;

Query Match 51.3%; Score 58; DB 7; Length 339;
 Best Local Similarity 38.5%; Pred. No. 1.78e+00;
 Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 276 QCVVVGKIDDF 288

QY 2 ECNIKVDVNDNF 14

RESULT 14
 ID Q12864 PRELIMINARY; PRT; 832 AA.
 AC Q12864;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE INTESTINAL PEPTIDE-ASSOCIATED TRANSPORTER HPT-1.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
 OC CATARRHINI; HOMINIDAE; HOMO.
 RN (1)

RP SEQUENCE FROM N.A.
 RC TISSUE-COLON;
 RX MEDLINE: 94204643.
 RA DANTZIG A.H., HOSKINS J., TABAS L.B., BRIGHT S., SHEPARD R.L.,
 RA JENKINS I.L., DUCKWORTH D.C., SPORTSMAN R., MACKENSEN D.,
 RA ROSTECK P.R., SKATRUD P.L.;
 RT "Association of intestinal peptide transport with a protein related
 to the cadherin superfamily";
 RL SCIENCE 264:430-433(1994).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: U07969; G483392; -;
 DR PROSITE: PS00232; CADHERIN; 3.
 DR PFAM: PF00028; cadherin; 6.
 KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
 SQ SEQUENCE 832 AA; 92147 MW; B5193483 CRC32;

Query Match 51.3%; Score 58; DB 4; Length 832;
 Best Local Similarity 64.3%; Pred. No. 1.78e+00;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 325 EIHVKVDINDNP 338

QY 2 ECNIKVDVNDNF 15

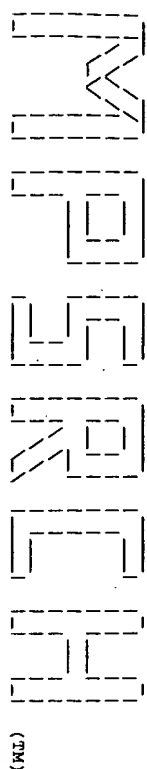
RESULT 15
 ID Q15336 PRELIMINARY; PRT; 832 AA.
 AC Q15336;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE LI-CADHERIN.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
 OC CATARRHINI; HOMINIDAE; HOMO.
 RN (1)
 RP SEQUENCE FROM N.A.
 RA BOETTINGER A., KREFT B., FIEGER C., DLOUHY B., BERNDORFF D.,
 RA GOESSNER R., TAUBER R.;
 RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: X83228; G854175; -;
 DR PROSITE: PS00232; CADHERIN; 3.
 DR PFAM: PF00028; cadherin; 6.
 KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.
 SQ SEQUENCE 832 AA; 92207 MW; 65C92D4C CRC32;

Query Match 51.3%; Score 58; DB 4; Length 832;
 Best Local Similarity 64.3%; Pred. No. 1.78e+00;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 325 EIHVKVDINDNP 338

QY 2 ECNIKVDVNDNF 15

Search completed: Fri Jun 11 17:38:42 1999
 Job time : 126 secs.



Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution Rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 11 17:45:41 1999; Maspar time 7.95 Seconds
40.130 Million cell updates/sec

Tabular output not generated.

Title: >US-08-991-628-6
Description: (1-15) from US08991628 pep
Perfect score: 109
Sequence: 1 SARTLNRRYTGPTTF 15

Scoring table:
PAM 150
Gap 15

Searched: 170751 seqs, 2126608 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

Database:

a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 19.152; Variance 58.684; scale 0.326

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	109	100.0	15 20	W04846	Self epitope of desmo	5.33e-05
2	109	100.0	15 20	W07908	Pemphigus vulgaris an	5.33e-05
3	109	100.0	999 6	R30742	Human pemphigus vulga	5.33e-05
4	56	51.4	480 2	R12098	Haem 84-2 portion of	6.15e+01
5	56	51.4	516 2	R12098	Haem 84-1 portion of	6.15e+01
6	55	50.5	431 17	R97293	Mouse CRF R81 recepto	7.81e+01
7	54	49.5	91 3	P61040	N-terminal region of	9.89e+01
8	54	49.5	246 8	R40924	Protein able to bind	9.89e+01
9	54	49.5	428 22	W08609	Chimeric MOMV and Ty	9.89e+01
10	54	49.5	428 22	W08609	Chimeric MOMV and Ty	9.89e+01
11	54	49.5	431 13	R90575	Rat CR2-beta recepto	9.89e+01
12	54	49.5	449 22	W08608	Chimeric MOMV and Ty	9.89e+01
13	54	49.5	1196 16	R75189	Osteoinductive retrov	9.89e+01
14	53	48.6	105 21	W00832	Variable light chain	1.25e+02
15	53	48.6	105 24	W19018	Anti-human FasL antib	1.25e+02
16	53	48.6	653 7	R37425	IDUA.	1.25e+02

17	53	48.6	2115 32	W59276	Rubella virus RA27/3	1.25e+02
18	53	48.6	2205 15	R79048	Infectious rubella v1	1.25e+02
19	52	47.7	105 29	W52241	Antibody LDI/2-6-3-V	1.58e+02
20	52	47.7	105 29	W52243	Antibody LDI/2-6-3-V	1.58e+02
21	52	47.7	454 17	R93616	Kaposi's sarcoma asso	1.58e+02
22	52	47.7	454 18	R97842	Kaposi's sarcoma asso	1.58e+02
23	52	47.7	691 2	R04711	Sequence of guinea pl	1.58e+02
24	52	47.7	1844 31	W56558	Toxin TcdA1, encoded	1.58e+02
25	52	47.7	1844 25	W18502	Photorehabdus luminesc	1.58e+02
26	52	47.7	2504 25	W17871	Photorehabdus luminesc	1.58e+02
27	52	47.7	2504 31	W56557	Toxin TcdA, encoded b	1.58e+02
28	51	46.8	39 5	R25101	bGFR prodng analogue	1.99e+02
29	51	46.8	45 5	R25254	bGFR prodng analogue	1.99e+02
30	51	46.8	106 8	R39567	Sequence of the varia	1.99e+02
31	51	46.8	107 8	R25413	Light chain variable	1.99e+02
32	51	46.8	108 5	R28751	Light chain variable	1.99e+02
33	51	46.8	127 20	W06447	Humc3 VL region BR-R	1.99e+02
34	51	46.8	128 36	W72435	Humanized V-kappa	1.99e+02
35	51	46.8	128 13	R70254	Transplanted V-kappa	1.99e+02
36	51	46.8	128 36	W72433	Transplanted V-kappa	1.99e+02
37	51	46.8	128 10	R55211	Humanized HPI/2 light	1.99e+02
38	51	46.8	129 13	R70256	Anti-VLA-4 humanized	1.99e+02
39	51	46.8	548 10	R54013	Human tissue transglu	1.99e+02
40	51	46.8	548 29	W46621	Amino acid sequence o	1.99e+02
41	51	46.8	687 6	R32074	Bovine transglutamina	1.99e+02
42	51	46.8	845 31	W56579	Fragment of toxin Tcd	1.99e+02
43	51	46.8	1306 34	W68155	Human angiotensin con	1.99e+02
44	51	46.8	1849 31	W56573	Toxin TcdA1, encoded	1.99e+02
45	51	46.8	2516 25	W17899	Photorehabdus luminesc	1.99e+02

ALIGNMENTS

RESULT 1
ID W04846 standard; peptide; 15 AA.
AC W04846:
DT 18-PFB-1997 (first entry)
DE Self epitope of desmoglein 3, implicated in autoimmune disease.
KW Tolerisation; self-epitope; antigen; autoimmune disease;
KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;
KW pemphigus vulgaris; desmoglein; multiple sclerosis;
KW herpes simplex virus; adenovirus; phosphomannomutase;
KW human papillomavirus; Epstein-Barr virus; DNA polymerase;
KW influenza; haemagglutinin; reovirus; sigma protein.
OS Homo sapiens.
PN W09627387-A1.
PS 12-SEP-1996.
PF 07-MAR-1996; U03182.
PR (HARD) HARVARD COLLEGE.
PI Strominger JL, Wucherpfennig KW;
DR WPI; 96-425218/42.
PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self
PT antigens - useful in disease treatment, and method for
PT identification of other self and non-self antigens implicated in
PT auto-immune disease
PS Claim 1, Page 41; 58pp; English.
CC Pharmaceutical preparations for tolerisation to antigens comprise
CC either an isolated human non-collagen or non-mysin basic protein
CC (MBP) polypeptide which is capable of tolerising an individual to an
CC autoantigen; or an isolated human pathogen polypeptide capable of
CC tolerising an individual to that polypeptide. In both cases, the
CC polypeptide (whether self or non-self) includes an amino acid
CC sequence corresponding to a sequence motif for a MHC class II
CC protein, such as HLA-DR, which is associated with a human autoimmune
CC disease and which binds to the polypeptide to activate autoreactive
CC T-cells in individuals with the autoimmune disease. This peptide is
CC derived from the human desmoglein 3 protein (amino acids 512-526)
CC and is implicated as a self epitope in pemphigus vulgaris. Peptides
CC derived from the human desmoglein protein are described in W04841-47.
SQ Sequence 15 AA;

Query Match 100.0%; Score 109; DB 20; Length 15;

Best Local Similarity 100.0%; Pred. No. 5.33e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 sartlnnrytgytff 15
OY 1 SARTLNRRYTGPTTF 15

RESULT 2

ID W07908 standard; protein; 614 AA.
AC W07908.1
DT 29-JAN-1997 (first entry)
DE Pemphigus vulgaris antigen protein extracellular region.
KW Autoantibody; immunoglobulin G; IgG1; fusion protein; diagnosis;
KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;
KW dermatology.
OS Homo sapiens.
PN J08188540-A.
PD 23-JUL-1996.
PR 30-JUN-1995; 165632.
PA (NISHU) NISHUKAWA T.
DR WPI: 96-388562/39.
PT Fused protein recognised by pemphigus vulgaris autoantibody -
PT useful to treat and diagnose pemphigus vulgaris
PS Claim 1; Page 7-9; 9PP; Japanese.
CC W07908 represents the human pemphigus vulgaris (PV) antigen
CC extracellular region. The PV antigen is produced in patients with
CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare
CC relapsing disease causing suprabasal, intra-epidermal bullae
CC (vesicles) of the skin and mucous membranes, which is fatal if
CC untreated. The PV antigen was fused to a human IgG1 hinge region
CC and the resulting fusion protein is useful to treat or diagnose
CC pemphigus vulgaris.
SQ Sequence 614 AA;

Query Match 100.0%; Score 109; DB 19; Length 614;
Best Local Similarity 100.0%; Pred. No. 5.33e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 511 sartlnnrytgytff 525
OY 1 SARTLNRRYTGPTTF 15

RESULT 3

ID R30742 standard; Protein; 999 AA.
AC R30742;
DT 14-JUN-1993 (first entry)
DE Human pemphigus vulgaris 130KD antigen.
KW Pemphigus vulgaris; skin disease; autoantibodies;
KW keratinocyte cell surface antigen; glycoprotein; cell adhesion.
OS Homo sapiens.
PN U57798918-A.
PD 15-DEC-1992.
PR 27-NOV-1991; 798918.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Amagel M, Klaus-Kovtun V, Stanley JR;
DR WPI: 93-067436/08.
PT N-PSDB; Q35992.
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
PT diagnostic and therapeutic uses
PS Disclosure: Fig 7; 50pp; English.
CC This sequence is the pemphigus vulgaris 130KD antigen. The protein
CC and its encoding DNA may be used in the diagnosis and treatment of
CC pemphigus vulgaris. It is thought that the antigen may be a cell
CC adhesion molecule.
SQ Sequence 999 AA;

Query Match 100.0%; Score 109; DB 6; Length 999;
Best Local Similarity 100.0%; Pred. No. 5.33e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 512 sartlnnrytgytff 526
OY 1 SARTLNRRYTGPTTF 15

RESULT 4

ID R12099 standard; Protein; 480 AA.
AC R12099;
DT 22-JUL-1991 (first entry)
DE Haem 84-2 portion of H.contortus 55A surface protein.
KW Parasitic nematode; vaccine; Hc.
OS Haemonchus contortus.
PN A09062569-A.
PD 21-MAR-1991.
PR 17-SEP-1990; 062569.
PR 18-SEP-1989; US-408339.
PA (SYNE-) SYNERGEN INC.
DR WPI: 91-133285/19.
DR N-PSDB; Q11799.
PT Protein from Haemonchus contortus and other nematodes - used as
PT therapeutic and prophylactic agent to protect plants, animals or
PT humans from parasitic nematode infection.
PS Disclosure: Fig 49; 209pp; English.
CC The proteins derived from the nematode DNA may be used to derive
CC vaccines against parasitic infection of plants, humans and animals
CC esp. sheep. Mabs may also be raised to provide passive therapy and
CC prophylaxis against infection.
SQ Sequence 480 AA;

Query Match 51.4%; Score 56; DB 2; Length 480;
Best Local Similarity 54.5%; Pred. No. 6.15e+01;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 376 vnnrygyrky 386
OY 5 LNNRRYTGPTTF 15

RESULT 5

ID R12098 standard; Protein; 516 AA.
AC R12098;
DT 22-JUL-1991 (first entry)
DE Haem 84-1 portion of H.contortus 55A surface protein.
KW Parasitic nematode; vaccine; Hc.
OS Haemonchus contortus.
PN A09062569-A.
PD 21-MAR-1991.
PR 17-SEP-1990; 062569.
PR 18-SEP-1989; US-408339.
PA (SYNE-) SYNERGEN INC.
DR WPI: 91-133285/19.
DR N-PSDB; Q11798.
PT Protein from Haemonchus contortus and other nematodes - used as
PT therapeutic and prophylactic agent to protect plants, animals or
PT humans from parasitic nematode infection.
PS Disclosure: Fig 49; 209pp; English.
CC The proteins derived from the nematode DNA may be used to derive
CC vaccines against parasitic infection of plants, humans and animals
CC esp. sheep. Mabs may also be raised to provide passive therapy and
CC prophylaxis against infection.
SQ Sequence 516 AA;

Query Match 51.4%; Score 56; DB 2; Length 516;
Best Local Similarity 54.5%; Pred. No. 6.15e+01;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 484 vnnrygyrky 494
OY 5 LNNRRYTGPTTF 15

OS	Chimeric - Moloney murine leukaemia virus.
OS	Chimeric - Saccharomyces cerevisiae transposon Ty3.
FT	Key
FT	Location/Qualifiers
FT	domain
FT	1..61
FT	/label= A
FT	/note= "From Ty3"
FT	62..283
FT	/label= B
FT	/note= "From MoMLV"
FT	284..428
FT	/label= C
FT	/note= "From MoMLV"
PN	MO9637626-A1.
PD	28-NOV-1996.
PF	10-MAY-1996: U06727.
PR	22-MAY-1995: US-445466.
PA	(CHIR) UNIV VIRGENE INC.
PA	(REGC) UNIV CALIFORNIA.
PI	Bilachone VM, Dillidine SL, Jolly DJ, Respass JG;
PI	Sandmeyer SB;
PI	WPI: 97-021329/02.
PT	New chimeric integrase for targeted vector integration in
PT	eukaryotic genomes - useful for gene therapy, providing more
PT	consistent gene expression and lower rates of insertional
PT	mutagenesis
PS	Claim 7: Page - : 98pp; English.
CC	The present sequence is a specific example of a chimeric integrase
CC	made up of three domains from MoMLV and Ty3, with at least one domain
CC	derived from Ty3 integrase. This protein can direct integration of a
CC	vector construct into a defined region of a target eukaryotic genome.
CC	As part of gene delivery vehicles and transduction competent
CC	recombinant retroviral particles it can be useful in somatic
CC	and germ cell gene therapy (in vivo or ex vivo) of a wide range of
CC	genetic, infectious, degenerative, inflammatory, cardiovascular and
CC	autoimmune diseases or cancer. Typical examples of the many potential
CC	applications include treatment of HIV infection, haemophilia, Alzheimer's
CC	disease and rheumatoid arthritis. This protein rather than wild-type
CC	retroviral integrase reduces the rate of insertional mutagenesis and the
CC	degree of variation in gene expression, and particularly does not disrupt
CC	tRNA gene expression.
CC	N.B. The present sequence is not shown in the specification, but is
CC	made up from the two original sequences as specified, MoMLV and Ty3,
CC	which are shown.
CC	Sequence 428 AA:
S0	
Query Match	49.5%; Score 54; DB 22; Length 428;
Best Local Similarity	46.2%; Pred. No. 9,89e+01;
Matches	6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
Db	360 tknlprwkgpyt 372
	:::1:1111
ty	2 ARTLNRYTGPT 14
RESULT	10
ID	W08606 standard; protein; 429 AA.
AC	W08606; 16-JUL-1997 (first entry)
DE	Chimeric MoMLV and Ty3 integrase designated AmbtCm.
KW	Moloney murine leukaemia virus; Saccharomyces cerevisiae;
KW	retrotransposon; yeast Ty3; position specific integration;
KW	inflammation; cardiovascular; autoimmune; cancer; HIV; haemophilia;
KW	Alzheimer's disease; rheumatoid arthritis; chimeric.
OS	Chimeric - Moloney murine leukaemia virus.
OS	Chimeric - Saccharomyces cerevisiae transposon Ty3.
FT	Key
FT	Location/Qualifiers
FT	domain
FT	1..41
FT	/label= A
FT	/note= "From MoMLV"
FT	42..284
FT	/label= B
FT	/note= "From Ty3"
FT	285..429
FT	domain

```

FT      /label= C
FT      /note= "From MOMVYL"
PD      MO9637626-A1.
PE      28-NOV-1996.
PF      10-MAY-1996; U06727.
PR      22-MAY-1995; US-445466.
PA      (CHIR ) CHIRON VIAGENE INC.
PB      (REGC ) UNIV CALIFORNIA.
PI      B1lachone VM, Dildine SL, Jolly DJ, Respass JG;
PI      Sandmeyer SB;
PI      WPI; 97-021229/02.
PT      New chimeric integrase for targeted vector integration in
PT      eukaryotic genomes - useful for gene therapy, providing more
PT      consistent gene expression and lower rates of insertional
PT      mutagenesis
PS      Claim 7; Page -: 98pp; English.
CC      The present sequence is a specific example of a chimeric integrase
CC      made up of three domains from MOMV and Ty3, with at least one domain
CC      derived from Ty3 integrase. This protein can direct integration of a
CC      vector construct into a defined region of a target eukaryotic genome.
CC      As part of gene delivery vehicles and transduction competent
CC      recombinant retroviral particles it can be useful in somatic
CC      and germ cell gene therapy (in vivo or ex vivo) of a wide range of
CC      genetic, infectious, degenerative, inflammatory, cardiovascular and
CC      autoimmune diseases or cancer. Typical examples of the many potential
CC      applications include treatment of HIV infection, hemophilia, Alzheimer's
CC      disease and rheumatoid arthritis. This protein rather than wild-type
CC      retroviral integrase reduces the rate of insertional mutagenesis and the
CC      degree of variation in gene expression, and particularly does not disrupt
CC      tRNA gene expression.
CC      N.B. The present sequence is not shown in the specification, but is
CC      made up from the two original sequences as specified, MOMV and Ty3,
CC      which are shown.
SQ      Sequence 429 AA;

Query Match      49.5%; Score 54; DB 22; Length 429;
Best Local Similarity 46.2%; Pied. No. 9,89e+01;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db      361 tknleprvbkpyt 373
      ::|::|::|
QY      2 ARTLNRRYRGPT 14

RESULT 11
ID      R90575 standard; Protein; 431 AA.
AC      R90575;
DT      08-APR-1996 (first entry)
DE      Rat GRF2-beta receptor.
KW      GRF2-beta receptor; corticotropin-releasing factor-2 receptor;
KW      cerebrovascular disorder; memory disorder; Alzheimer disease.
OS      Rattus sp.
FH      Key
FH      domain
FT      1..117 Location/Qualifiers
FT      /label= Extracellular_N-terminal_domain
FT      118..138
FT      /label= Transmembrane_domain
FT      139..147
FT      /label= Intracellular_domain
FT      148..167
FT      /label= Transmembrane_domain
FT      168..184
FT      /label= Extracellular_domain
FT      185..208
FT      /label= Transmembrane_domain
FT      209..223
FT      /label= Intracellular_domain
FT      224..244
FT      /label= Transmembrane_domain
FT      245..261
FT      /label= Extracellular_domain
FT      262..286
FT      /label= Transmembrane_domain

```

```

FT domain 287..309 /Label= Intracellular_domain
FT domain 310..329 /Label= Transmembrane_domain
FT domain 330..342 /Label= Transmembrane_domain
FT domain 343..363 /Label= Extracellular_domain
FT domain 364..431 /Label= Transmembrane_domain
FT domain /Label= C-terminal_intracellular_domain
PN W0934651-A2.
PD 21-DEC-1995.
PF 14-JUN-1995; U07757.
PR 14-JUN-1994; US-2595959.
PR 31-JAN-1995; US-381433.
PR 07-JUN-1995; US-485984.
PA (NEUR-) NEUROCRINE BIOSCIENCES INC.
PI Chalmers D, De Souza EB, Grigoriadis DE, Liaw CM,
PI Lovenberg TW, Oltersdorf T;
PI WPI; 96-049680/05.
DR N-PDB; T12244.
PT Coricotropin-releasing factor-2 receptor, and DNA encoding it
PT used to isolate CRF-2 receptor antagonists for the treatment of
PT neurovascular disorders, memory disorders and Alzheimer's disease
PS Disclosure; Page 63-66; 109pp; English.
CC Rat corticotropin-releasing factor-2-beta (CRF2-beta) receptor
CC (R90575) is a membrane-bound G-coupled protein receptor involved
CC in signal transduction. It can be produced by expression of
CC encoding cDNA (T12244) in prokaryotic or eucaryotic host cells.
CC Recombinant CRF2 receptor is used to screen CRF2 receptor agonists
CC and antagonists of therapeutic appln., and to prepare antibodies
CC which specifically bind to CRF2 receptors.
SQ Sequence 431 AA:

Query Match 49.5%; Score 54; DB 15; Length 431;
Best Local Similarity 38.5%; Pred. No. 9.89e+01;
Matches 5; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

DB 47 rtttrfsgpyy 59
   ||::|||::
OY 3 RRLNNRYTGPT 15

RESULT 12
ID W08608 standard; Protein: 449 AA.
AC W08608:
DE 16-JUL-1997 (first entry)
DE Chimeric MoMLV and Ty3 integrase designated AtbCm.
KM Moloney murine leukemia virus; Saccharomyces cerevisiae;
KM retrotransposon; yeast Ty3; position specific integration;
KM inflammation; cardiovascular; autoimmune; cancer; HIV; haemophilia;
KM Alzheimer's disease; rheumatoid arthritis; chimeric.
OS Chimeric - Moloney murine leukemia virus.
OS Chimeric - Saccharomyces cerevisiae transposon Ty3.
FH Key Location/Qualifiers
FT domain 1..61 /Label= A
FT domain /Label= "From Ty3"
FT domain 62..304 /Label= B
FT domain /Label= "From Ty3"
FT domain 305..449 /Label= C
FT domain /Label= "From MoMLV"
PN W09637626-A1.
PD 28-NOV-1996.
PF 10-MAY-1996; U06727.
PR 22-MAY-1995; US-445466.
PA (CHIR) CHIRON VIRGENE INC.
PA (REGC) UNIV CALIFORNIA.
PI Blichone VM, Dildine SL, Jolly DJ, Respass JG;
PI Sandmeyer SB;
PI WPI; 97-021229/02.

```

```

PT New chimeric integrase for targeted vector integration in
PT eukaryotic genomes - useful for gene therapy, providing more
PT consistent gene expression and lower rates of insertional
PT mutagenesis
PS Claim 7; Page 7; 98pp; English.
CC The present sequence is a specific example of a chimeric integrase
CC made up of three domains from MoMLV and Ty3, with at least one domain
CC derived from Ty3 integrase. This protein can direct integration of a
CC vector construct into a defined region of a target eukaryotic genome.
CC As part of gene delivery vehicles and transduction competent
CC recombinant retroviral particles it can be useful in somatic
CC and germ cell gene therapy (in vivo or ex vivo) of a wide range of
CC genetic, infectious, degenerative, inflammatory, cardiovascular and
CC autoimmune diseases or cancer. Typical examples of the many potential
CC applications include treatment of HIV infection, haemophilia, Alzheimer's
CC disease and rheumatoid arthritis. This protein rather than wild-type
CC retroviral integrase reduces the rate of insertional mutagenesis and the
CC degree of variation in gene expression, and particularly does not disrupt
CC rRNA gene expression.
CC N.B. The present sequence is not shown in the specification, but is
CC made up from the two original sequences as specified, MoMLV and Ty3,
CC which are shown.
SQ Sequence 449 AA:

Query Match 49.5%; Score 54; DB 22; Length 449;
Best Local Similarity 46.2%; Pred. No. 9.89e+01;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

DB 381 tknlprwkypt 393
   ::::|::|::|
OY 2 ARLNNRYTGPT 14

RESULT 13
ID R75189 standard; Protein: 1196 AA.
AC R75189;
DE 30-MAY-1996 (first entry)
DE Osteoinductive retrovirus RFB-14 pol gene product.
DE RFB retrovirus; gag; pol; env; osteogenesis; osteoinductive protein;
KM bone development; osteoporosis; gene therapy; polymerase;
KM reverse transcriptase.
OS Retrovirus RFB-14.
OS DE4411718-A1.
PN DE4411718-A1.
PD 12-OCT-1995.
PF 05-APR-1994; 411718.
PR 05-APR-1994; DE-411718.
PA (GSFV-) GSF FORSCHUNGSZENTRUM UMWEILT & GESUNDHEIT.
PI Erdle V, Gimbel W, Oestergaard W, Pedersen FS, Pedersen L;
PI Schmidt J, Straus P;
PI WPI; 95-352078/46.
DR N-PSDB; 094266.
PT RFB-14 retrovirus genome - and prodn. of osteo-inductive proteins
PS Claim 14; Fig 1; 46pp; German.
CC The full-length proviral genomic sequence of retrovirus RFB-14 has
CC been determined. The virus codes for an osteoinductive protein,
CC although the precise location of the coding region has not yet been
CC identified. The virus may be useful in gene therapy of bone growth
CC disorders such as osteoporosis. The present sequence is that of the
CC viral pol gene product.
SQ Sequence 1196 AA:

Query Match 49.5%; Score 54; DB 16; Length 1196;
Best Local Similarity 46.2%; Pred. No. 9.89e+01;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

DB 1131 tknlprwkypt 1143
   ::::|::|::|
OY 2 ARLNNRYTGPT 14

RESULT 14
ID W00832 standard; Protein: 105 AA.
AC W00832;

```

29-MAY-1997 (first entry)
 DE Variable light chain of anti-human Fas ligand antibody NOK-5.
 KW Variable region; light chain; human; Fas ligand; monoclonal;
 KW antibody; NOK-5; hybridoma; inhibition; apoptosis; assay;
 KW diagnosis; disease; hepatitis; infectious mononucleosis;
 KW systemic lupus erythematosus.
 OS Mus musculus.
 PN WO9629350-A1.
 PD 26-SEP-1996.
 PR 21-MAR-1996; J00734.
 PR 20-MAR-1995; JP-087420.
 PR 27-OCT-1995; JP-303492.
 RA (SUME) SUMITOMO ELECTRIC IND CO.
 PI Kayagaki N, Nakata M, Okumura K, Yagita H;
 DR N-PSDB: T39558.
 PT Monoclonal antibody specifically recognising the Fas ligand - useful
 PT for the detection of Fas ligands either on cell surface or in
 PI solution
 PS Claim 28; Page 91; 133pp; Japanese.
 CC The present sequence is the light chain variable region of the
 CC anti-human Fas ligand monoclonal antibody (Mab) NOK-5. NOK-5 is
 CC produced by the hybridoma NOK-5 (FERM BP-5048), which was prepared
 CC by immunising mice with transformed human Fas ligand expressing COS
 CC cells, and fusing spleen cells isolated from the mice with myeloma
 CC P3x63Ag8.653 (ATCC CRL-1580) cells. The Mab recognises the human
 CC Fas ligand on the cell surface or in solution, and can be used to
 CC inhibit the apoptosis inducing cell surface Fas ligand/Fas
 CC reaction. The Mab can also be used for a Fas ligand assay in
 CC biological samples (e.g. human blood), especially for disease
 CC diagnosis, e.g. hepatitis, infectious mononucleosis and systemic
 CC lupus erythematosus.
 SQ Sequence 105 AA;

Query Match 48.6%; Score 53; DB 21; Length 105;
 Best Local Similarity 62.5%; Pred. No. 1.25e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

DB 91 hysspytf 98
 :|::|||
 QY 8 RYTGPTTF 15

RESULT 15
 ID W19018 standard; Protein; 105 AA.
 AC W19018;
 DT 14-JAN-1998 (first entry)
 DE Anti-human FasL antibody (NOK5) light chain variable region.
 KW Light chain; variable region; mouse; murine; human; Fas ligand;
 KW FasL; monoclonal antibody; Mab; hybridoma; treatment; hepatitis;
 KW hepatitis B virus; HBV; hepatitis C virus; HCV; apoptosis;
 KW liver cell; glutamate oxaloacetate; pyruvate transaminase.
 OS Mus sp.
 PN MO9715326-A1.
 PD 01-MAY-1997.
 PR 24-OCT-1996; J03089.
 PR 27-OCT-1995; JP-303491.
 PA (SUME) SUMITOMO ELECTRIC IND CO.
 PI Kayagaki N, Nakata M, Okumura K, Seino K, Yagita H;
 DR WPI: 97-258767/23.
 DR N-PSDB: T69542.
 PT Anti-human Fas Ligand antibody to treat hepatitis - controls
 PT apoptosis in liver cells and improves liver function
 PS Claim 6; Page 41; 51pp; Japanese.
 CC The present sequence is the light chain variable region of the
 CC murine anti-human Fas ligand (FasL) monoclonal antibody (Mab) NOK5,
 CC which is expressed by the hybridoma NOK5 (FERM BP-5044). The Mab
 CC can be used in the preparation of a composition for the effective
 CC oral or parenteral treatment of hepatitis, including hepatitis
 CC caused by hepatitis B or C virus. The composition controls apoptosis
 CC in liver cells caused by the binding of FasL to Fas expressing liver
 CC cells, and improves liver function by improving blood glutamate
 CC oxaloacetate and pyruvate transaminase levels. The composition is

CC given in a dosage of 0.0001-1000, preferably 0.01-600 mg/day.
 CC Spleen cells from mice immunised with FasL expressing COS cells
 CC were fused with mouse myeloma cells to produce hybridomas. The
 CC hybridomas were screened for anti-FasL activity, and the active
 CC clones NOK1-5 isolated.
 SQ Sequence 105 AA;

Query Match 48.6%; Score 53; DB 24; Length 105;
 Best Local Similarity 62.5%; Pred. No. 1.25e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

DB 91 hysspytf 98
 :|::|||
 QY 8 RYTGPTTF 15

Search completed: Fri Jun 11 17:50:17 1999
 Job time : 276 secs.

Db 512 SARTLNRYTGPYTF 526
QY 1 SARTLNRYTGPYTF 15

RESULT 2
ENTRY 2
TITLE S25196 #type complete
mitochondrial import protein MP1 precursor - yeast
(Saccharomyces cerevisiae)
ALTERNATE_NAMES protein Y13299.09; protein Y11022w
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 28-May-1993 #sequence_revision 28-May-1993 #text_change
12-Dec-1997

ACCESSIONS S25196; S49960; S35357
REFERENCE S25196
#authors Maarse, A.C.; Blom, J.; Grievell, L.A.; Meljer, M.
#journal EMBO J. (1992) 11:3619-3628
#title MP1, an essential gene encoding a mitochondrial membrane
protein, is possibly involved in protein import into yeast
mitochondria.
#cross-references M01D:93010953
#accession S25196
#molecule_type DNA
#residues 1-431 #label MAX
#cross-references EMBL:X67276; NID:g3966; PID:g3967
S49951
REFERENCE S49951
#authors Skelton, J.; Churcher, C.
#submission submitted to the EMBL data library, December 1994
#accession S49960
#molecule_type DNA
#residues 1-431 #label SKE
#cross-references EMBL:246881; NID:g599967; PID:g599977; MIPS:Y11022w
S35357
REFERENCE S35357
#authors Horst, M.; Jenoe, P.; Kronidou, N.G.; Bolliger, L.; Oppiger,
W.; Scherer, P.; Manning-Krieg, U.; Jascut, T.; Schatz, G.
#journal EMBO J. (1993) 12:3035-3041
#title Protein import into yeast mitochondria: the inner membrane
import site protein ISP45 is the MP1 gene product.
#accession S35357
#molecule_type protein
#residues 82-89;112-122;221-228;295-306;316-327;356-370 #label
HOR

GENETICS SGD:TIM44; MP1
#gene #cross-references MIPS:Y11022w; SGD:S0001284
#map_position 9L
#genome nuclear
KEYWORDS membrane protein; mitochondrion
SUMMARY #length 431 #molecular_weight 48854 #checksum 2985

Query Match 59.6%; Score 65; DB 2; Length 431;
Best Local Similarity 57.1%; Pred. No. 2.40e+01;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 13 SSRILNRYTGPYTF 26
QY 1 SARTLNRYTGPYTF 14

RESULT 3
ENTRY 3
TITLE T00045 #type complete
cellodextrin phosphorylase - Clostridium thermocellum
ORGANISM #formal_name Clostridium thermocellum
DATE 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
22-Jan-1999

ACCESSIONS T00045
REFERENCE T00045
#authors Kawaguchi, T.; Ikeuchi, Y.; Tsutsumi, N.; Kan, A.; Sumitani,
J.; Arai, M.
#journal J. Ferment. Bioeng. (1998) 85:144-149
#title Cloning, nucleotide sequence, and expression of the
Clostridium thermocellum cellodextrin phosphorylase gene

and its application to synthesis of cellulase inhibitors.
#accession T00045
#status Preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-980 #label KAW
#cross-references EMBL:AB006822; NID:d117395; PID:d1022940
#experimental_source ATCC 27405

GENETICS
SUMMARY #cdp
#length 980 #molecular_weight 111183 #checksum 2409

Query Match 56.0%; Score 61; DB 2; Length 980;
Best Local Similarity 60.0%; Pred. No. 1.30e+00;
Matches 9; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

Db 713 ARVLNRYKDSYTF 727
QY 2 ARTLNRYTGPYTF 15

RESULT 4
ENTRY 4
TITLE S43577 #type complete
C28A5.4 protein (clone C28A5) - Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 07-Sep-1994 #sequence_revision 10-Nov-1995 #text_change
17-Oct-1997

ACCESSIONS S43577
REFERENCE S43577
#authors Palmer, S.
#submission submitted to the EMBL Data Library, April 1994
#accession S43577
#status Preliminary
#molecule_type DNA
#residues 1-282 #label PAL
#cross-references EMBL:232680

GENETICS
#introns 77/1; 120/2; 145/3; 225/3
CLASSIFICATION #superfamily unassigned homeobox proteins; homeobox homology
KEYWORDS DNA binding; homeobox; nucleus; transcription regulation
FEATURE 103-168
SUMMARY #domain homeobox homology #status atypical #label HOX

Query Match 55.0%; Score 60; DB 2; Length 282;
Best Local Similarity 40.0%; Pred. No. 2.07e+00;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 224 PPMTLNRYTGPYTF 238
QY 1 SARTLNRYTGPYTF 15

RESULT 5
ENTRY 5
TITLE G8BP74 #type complete
baseplate protein gp8 - phage T4
ALTERNATE_NAMES baseplate wedge protein
ORGANISM #formal_name phage T4
#note host Escherichia coli
DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
05-Sep-1997

ACCESSIONS J00658; JF0058
REFERENCE J00656
#authors Efimov, V.P.; Prilipov, A.G.; Mesyanzhinov, V.V.
#journal Nucleic Acids Res. (1990) 18:5313
#title Nucleotide sequences of bacteriophage T4 genes 6, 7 and 8.
#cross-references M01D:90384864
#accession J00658
#molecule_type DNA
#residues 1-334 #label EFI
#cross-references EMBL:X15907; NID:g15321; PID:g15324
JF0058
REFERENCE JF0058
#authors Mesyanzhinov, V.V.
#submission submitted to JFID, December 1989

#accession JF0058
#molecule_type DNA
#residues 1-334 #label MES
COMMENT The order of association in the morphogenesis pathway is strictly determined except for gp 11 which can be added at any stage of the pathway. GP 8 associates to the precursor structure after gp 7 binding and before gp 6 binding.

GENETICS
#gene gp8
#map_position 85,20-86,20
#note gene 8 is directed clockwise in the T4 map; the 5'-end of gene 8 overlaps with the 3'-end of gene 7
#superfamily phage T4 baseplate protein gp8
#baseplate: late protein
#length 334 #molecular-weight 38007 #checksum 5888

CLASSIFICATION
SUMMARY

Query Match 55.0%; Score 60; DB 1; Length 334;
Best Local Similarity 60.0%; Pred. No. 2.07e+00;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 97 DRRPDPYTF 106
QY 6 NNRYTGPYTF 15

RESULT 6
ENTRY T01862 #type complete
TITLE hypothetical protein T7M24.4 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear cress
DATE 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 24-Mar-1999

ACCESSIONS
REFERENCE T01862
#authors Harmon, G.; Langston, Y.; Stoneking, T.; Drone, K.; Ames, M.
#submission submitted to the EMBL Data Library, July 1998
#description The sequence of Arabidopsis thaliana T7M24.
#accession T01862
#status translated from GB/EMBL/DDBJ
#molecule_type DNA
#residues 1-973 #label HAR
#cross-references EMBL:AF077408; NID:g3319363
#experimental_source cultivar Columbia

GENETICS
#map_position 4
#introns 257/1; 284/1; 378/3; 689/3; 744/3
#note T7M24.4
SUMMARY #length 973 #molecular-weight 112369 #checksum 1656

Query Match 55.0%; Score 60; DB 2; Length 973;
Best Local Similarity 53.8%; Pred. No. 2.07e+00;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 826 SRKISPRVGPY 838
QY 1 SARILNNRYTGPY 13

RESULT 7
ENTRY B70854 #type complete
TITLE hypothetical protein Rv2991 - Mycobacterium tuberculosis (strain H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
ACCESSIONS B70854
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry, C.E.; Tekale, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;

Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
#cross-references MIMD:98295987
#accession B70854
#status preliminary; nucleic acid sequence not shown; translation not shown

GENETICS
#gene Rv2991
#map_position 1-163 #label COL
#residues 1-163 #label COL
#cross-references GB:A021287; GB:AL123456; NID:g3261508; PID:41237736; PID:g2791588
#experimental_source strain H37Rv

CLASSIFICATION
SUMMARY

Query Match 54.1%; Score 59; DB 2; Length 163;
Best Local Similarity 87.5%; Pred. No. 3.15e+00;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 111 ERYTGPYTF 118
QY 7 NNRYTGPYTF 14

RESULT 8
ENTRY C71180 #type complete
TITLE Probable N-acetyl-gamma-glutamyl-phosphate reductase - Pyrococcus horikoshii
ORGANISM #formal_name Pyrococcus horikoshii
DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
ACCESSIONS C71180
REFERENCE A71000
#authors Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Koki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuch, Y.; Shizuya, H.; Kikuchi, H.
#journal DNA Res. (1998) 5:55-76
#title Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3
#cross-references MIMD:98344137
#accession C71180
#status preliminary; nucleic acid sequence not shown; translation not shown

GENETICS
#gene PH1720
#map_position 1-330 #label KAM
#residues 1-330 #label KAM
#cross-references GB:AP000007; NID:g3236134; PID:41031777; PID:g3258151
#experimental_source strain OT3
#note this accession replaces an interim accession for a sequence replaced by GenBank

CLASSIFICATION
SUMMARY

Query Match 54.1%; Score 59; DB 2; Length 330;
Best Local Similarity 53.8%; Pred. No. 3.15e+00;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 50 RGLNRTSYDNF 62
QY 3 RLNNRYTGPYTF 15

RESULT 9

ENTRY S6088 #type complete
TITLE ferric exochelin biosynthesis protein fxbA - Mycobacterium
ORGANISM #formal_name Mycobacterium smegmatis
DATE 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change
17-Mar-1999

ACCESSIONS S6088
REFERENCE S6088
#authors Fiss, E.H.; Yu, S.; Jacobs Jr., W.R.
#journal Mol. Microbiol. (1994) 14:557-569
#title Identification of genes involved in the sequestration of iron
in mycobacteria: the ferric exochelin biosynthetic and
uptake pathways.
#cross-references MIMD:95191405
#accession S6088
#molecule_type DNA
#residues 1-360 #label FIS
#cross-references EMBL:U10425; NID:g595400; PID:g595404

GENETICS
#gene fxbA
CLASSIFICATION #superfamily methionyl-tRNA formyltransferase:
phosphoribosylglycinamide formyltransferase homology
FEATURE 20-209 #domain phosphoribosylglycinamide formyltransferase
#homology #label PRGF
SUMMARY #length 360 #molecular-weight 41130 #checksum 7127

Query Match 54.1%: Score 59; DB 2; Length 360;
Best Local Similarity 38.5%: Pred. No. 3.15e+00;
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 241 RAISDPYPNATY 253
|||:|:|:
QY 3 RLNNRYTGPYTF 15

RESULT 10
ENTRY C70333 #type complete
TITLE hypothetical protein aq_372 - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change
08-May-1998

ACCESSIONS C70333
REFERENCE A70300
#authors Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;
Keller, M.; Aulay, M.; Huber, R.; Feldman, R.A.; Short,
J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium
Aquifex aeolicus.
#cross-references MIMD:98196666
#accession C70333
#status preliminary: nucleic acid sequence not shown;
translation not shown

GENETICS
#molecule_type DNA
#residues 1-531 #label AQF
#cross-references GB:AE000686; NID:g2983038; PID:g2983043; GB:AE000657
#experimental_source strain VF5

SUMMARY #gene aq_372
#length 531 #molecular-weight 62830 #checksum 4972

Query Match 54.1%: Score 59; DB 2; Length 531;
Best Local Similarity 66.7%: Pred. No. 3.15e+00;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 80 TLKRLNPYTF 91
|||:|:|:
QY 4 TLNNRYTGPYTF 15

RESULT 11

ENTRY E69312 #type complete
TITLE nitrate reductase gamma subunit homolog - Archaeoglobus
fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
05-Jun-1998

ACCESSIONS E69312
REFERENCE A69250
#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.;
Peterson, J.D.; Richardson, D.L.; Kervilavage, A.R.; Graham,
D.E.; Kyriides, N.C.; Fleischmann, R.D.; Quackenbush, J.;
Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.;
Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman,
J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,
T.; Artlich, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;
D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
C.R.; Venter, J.C.
#journal Nature (1997) 390:364-370
#title The complete genome sequence of the hyperthermophilic,
sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references MIMD:98049343
#accession E69312
#status preliminary: nucleic acid sequence not shown;
translation not shown

GENETICS
#molecule_type DNA
#residues 1-332 #label KLE
#cross-references GB:AE001069; GB:AE000782; NID:g2689392; PID:g2650124;
TIGR:AF0501

SUMMARY #length 332 #molecular-weight 38386 #checksum 9937

Query Match 53.2%: Score 58; DB 2; Length 332;
Best Local Similarity 50.0%: Pred. No. 4.77e+00;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 53 RTIYDRFDSPT 64
|||:|:|:
QY 3 RLNNRYTGPYTF 14

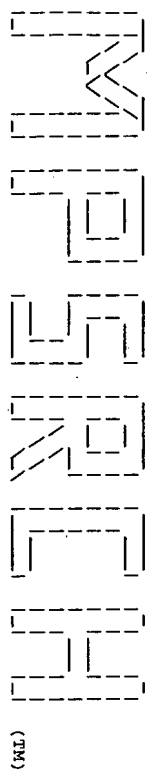
RESULT 12
ENTRY F69075 #type complete
TITLE nltrogenase (EC 1.18.6.1) molybdenum-iron protein alpha chain
ORGANISM #formal_name Methanobacterium thermoautotrophicum
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
16-Dec-1998

ACCESSIONS F69075
REFERENCE A69000
#authors Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.;
Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.;
Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keegle, P.;
Lunn, W.; Potlher, B.; Qiu, D.; Spadafora, R.; Vicario, R.;
Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso,
A.; Bush, D.; Saefer, H.; Patwell, D.; Prabhakar, S.;
McDougal, S.; Shimer, G.; Goyal, A.; Pietrowski, S.;
Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
J.; Reeve, J.N.
#journal J. Bacteriol. (1997) 179:7135-7155
#title Complete genome sequence of Methanobacterium
thermoautotrophicum Delta H: functional analysis and
comparative genomics.
#cross-references MIMD:98037514
#accession F69075
#status preliminary: nucleic acid sequence not shown;
translation not shown

GENETICS
#molecule_type DNA
#residues 1-469 #label MTH
#cross-references GB:AE000916; GB:AE000666; NID:g2622674; PID:g2622665
#experimental_source strain Delta H

GENETICS
#gene MTH1563
CLASSIFICATION #superfamily dinitrogenase alpha chain; nitrogenase
#vanadium-iron protein alpha chain homology
KEYWORDS oxidoreductase
FEATURE
1-469
#domain nitrogenase vanadium-iron protein alpha chain
SUMMARY
#homology #label VIA
#length 469 #molecular-weight 52866 #checksum 1591
Query Match 53.2%; Score 58; DB 2; Length 469;
Best Local Similarity 58.3%; Pred. No. 4,77e+00;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
DB 273 ARLEARNVNPY 284
OY 2 ARTLNRRYTGPT 13
RESULT 13
ENTRY PNO105 #type complete
TITLE RNA-directed RNA polymerase (EC 2.7.7.48) 85K protein -
barley stripe mosaic virus (strain 4-2)
ALTERNATE_NAMES strain 12-2; strain 7-2
ORGANISM #formal_name Barley stripe mosaic virus, BSMV
#accession PNO105; PNO102; PNO107; PNO106
#date 29-May-1998
#text_change 30-Sep-1993 #sequence_revision 30-Sep-1993
REFERENCE
#authors Kozlov, Y.V.; Afanasiev, B.N.; Rupasov, V.V.; Golova, Y.B.;
Kulaveva, O.I.; Dolja, V.V.; Atabekov, J.G.; Bayev, A.A.
#journal Mol. Biol. (Mosk.) (1989) 23:1080-1090
#title The complete nucleotide sequence of barley stripe mosaic
virus RNA 3 and its variability.
#cross-references MUID:9006400
#accession PNO105
#molecule_type genomic RNA
#residues 1-770 #label KOZ
#experimental_source strain 4-2
#accession PNO102
#molecule_type genomic RNA
#residues 122-770 #label KOZ
#accession PNO107
#molecule_type genomic RNA
#residues 1-118,120-770 #label KO3
#experimental_source strain 12-2
#accession PNO106
#molecule_type genomic RNA
#residues 1-118, 125-770 #label KO4
#experimental_source strain 7-2
COMMENT the duplication in this protein results from a tandem duplication
in the virus genome found in some, but not all, virus strains.
GENETICS
#map_position segment 3(II)
KEYWORDS duplication; nucleotidyltransferase
FEATURE
1-118
#region duplication
SUMMARY
#length 770 #molecular-weight 87176 #checksum 8385
Query Match 53.2%; Score 58; DB 2; Length 770;
Best Local Similarity 50.0%; Pred. No. 4,77e+00;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
DB 720 KALVNRYSPPYS 731
OY 3 RTLNRRYTGPT 14
RESULT 14
ENTRY GNEFA2 #type complete
TITLE retrovirus-related pol polyprotein - fruit fly (Drosophila
melanogaster). retrotransposon 412

CONTAINS
#formal_name retropepsin (EC 3.4.23.16)
#formal_name Drosophila melanogaster
DATE 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change
22-Jan-1999
ACCESSIONS D29349
REFERENCE A91171
#authors Yuki, S.; Inoue, S.; Ishimaru, S.; Satgo, K.
#journal Eur. J. Biochem. (1986) 158:403-410
#title Nucleotide sequence characterization of a Drosophila
retrotransposon, 412.
#cross-references MUID:86274717
#accession D29349
#molecule_type DNA
#residues 1-1237 #label YUK
#cross-references GB:X04132; GB:X03733; NID:98500; PID:e1912;
PID:g1335652
#note the authors translated the codon CAA for residue 21 as
Lys
GENETICS
#gene FlyBase:412
CLASSIFICATION #cross-references FlyBase:Fg9n000006
#superfamily pol polyprotein
KEYWORDS aspartic proteinase; hydrolase; polypeptide; reverse
transcriptase
FEATURE
39-133
#product retropepsin #status predicted #label RTP\
#active_site Asp (shared with dimeric partner) #status
predicted
SUMMARY
#length 1237 #molecular-weight 143041 #checksum 6456
Query Match 53.2%; Score 58; DB 1; Length 1237;
Best Local Similarity 42.9%; Pred. No. 4,77e+00;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
DB 1193 GHRKDFKTYGPKI 1206
OY 2 ARTLNRRYTGPT 15
RESULT 15
ENTRY S48546 #type complete
TITLE hypothetical protein YLR193c - yeast (Saccharomyces
cerevisiae)
ALTERNATE_NAMES hypothetical protein L8167.12
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change
12-Dec-1997
ACCESSIONS S48546
REFERENCE S48545
#authors Pauley, A.
#submissions submitted to the EMBL Data Library, September 1994
#description The sequence of S. cerevisiae cosmid 8167.
#accession S48546
#molecule_type DNA
#residues 1-175 #label PAU
#cross-references EMBL:U14913; NID:9544497; PID:9544509; MIPS:YLR193c
GENETICS
#map_position 12R
SUMMARY
#length 175 #molecular-weight 20108 #checksum 5250
Query Match 52.3%; Score 57; DB 2; Length 175;
Best Local Similarity 46.2%; Pred. No. 7,17e+00;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
DB 19 SRAFNRYNPYS 31
OY 2 ARTLNRRYTGPT 14
Search completed: Fri Jun 11 17:52:05 1999
Job time : 92 secs.



Release 3.1A John F. Collins, Bioinformatics Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPearch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Jun 11 17:45:15 1999; MasPar time 2.45 Seconds
173.166 Million cell updates/sec
Tabular output not generated.

Title: >US-08-991-628-6
Description: (1-15) from US08991628.pep
Perfect Score: 109
Sequence: 1 SARFLNNRYGTPTF 15

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 27.195; Variance 33.825; scale 0.804

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	109	100.0	999	1	DSG3_HUMAN	DESMOGLEIN 3 PRECURSOR	8.99e+13
2	65	59.6	431	1	IM44_YEAST	MITOCHONDRIAL IMPORT I	5.37e+02
3	62	56.0	391	1	GA6B_XENLA	TRANSFERRIN FACTOR X	2.30e+01
4	60	55.0	273	1	HMD1_CAEL	PUTATIVE HOMODOM PROT	5.92e+01
5	60	55.0	334	1	VG08_BP74	BASEPLATE STRUCTURAL P	5.92e+01
6	59	54.1	304	1	CBP2_SIMV1	ZINC CARBOXYPEPTIDASE	9.43e+01
7	58	53.2	469	1	NIFD_METH	NITROGENASE MOLYBDENUM	1.49e+00
8	58	53.2	1237	1	POL4_DROME	RETROVIRUS-RELATED POL	1.49e+00
9	57	52.3	387	1	GAT6_CHICK	TRANSFERRIN FACTOR G	2.35e+00
10	57	52.3	1013	1	CHS2_EMENT	CHITIN SYNTHASE 2 (EC	2.35e+00
11	56	51.4	360	1	IDH1_YEAST	ISOCITRATE DEHYDROGENA	3.68e+00
12	56	51.4	1317	1	N145_YEAST	NUCLEOPORIN NUP145 (NU	3.68e+00
13	55	50.5	245	1	GYPL_ANAPL	HYPOPHOSPHATE 28.3 KD P	5.72e+00
14	55	50.5	391	1	VAZC_RHNS	HYPOPHOSPHATE 28.3 KD P	5.72e+00
15	55	50.5	391	1	GA6A_XENLA	TRANSFERRIN FACTOR X	5.72e+00
16	55	50.5	431	1	CRF2_MOUSE	CORTICOTROPIN RELEASIN	5.72e+00
17	55	50.5	433	1	CBPA_ANOGA	ZINC CARBOXYPEPTIDASE	5.72e+00
18	55	50.5	455	1	KR2_EBV	PROBABLE SERINE/THREON	8.85e+00
19	54	49.5	120	1	POLI_MOUSE	RETROVIRUS-RELATED POL	8.85e+00
20	54	49.5	130	1	V29K_TRYNC	29 KD PROTEIN (FRAGMEN	8.85e+00
21	54	49.5	194	1	VA43_YACCV	PROTEIN A4.3	8.85e+00
22	54	49.5	231	1	RECO_COXBU	DNA REPAIR PROTEIN REC	8.85e+00
23	54	49.5	252	1	V29K_TRYSV	29 KD PROTEIN.	8.85e+00

ALIGNMENTS

RESULT ID	1	DSG3_HUMAN	STANDARD:	PRT:	999 AA.
AC	P32926;				
DT	01-OCT-1993 (REL. 27, CREATED)				
DT	01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)				
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)				
DE	DESMOGLEIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).				
GN	DSG3				
OS	HOMO SAPIENS (HUMAN).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;				
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE: 92069753.				
RA	AMAGAI M., KLAUS-KOVYUN V., STANLEY J.R.;				
RT	"Autoantibodies against a novel epithelial cadherin in pemphigus				
RT	vulgaris, a disease of cell adhesion."				
RL	CELL 67:869-877(1991).				
CC	- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSE JUNCTIONS.				
CC	- INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND IMMEDIATE				
CC	FILAMENTS MEDIATING CELL-CELL ADHESION.				
CC	- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.				
CC	- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, OESOPHAGUS AND				
CC	CARCINOMAS.				
CC	- DOMAIN: CALCIN MAY BE BOUND BY THE CADHERIN-LIKE REPEATS				
CC	(POTENTIAL).				
CC	- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN				
CC	DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE				
CC	LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES				
CC	AGAINST DSG3.				
CC	- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE				
CC	DESMOSOMAL SUBFAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL: M76482; GI90752; -				
CC	PIR: A14088; IJH03.				
CC	MTM: 169615; -				
CC	PROSITE: PS00227; CADHERIN; 3.				
CC	PFAM: PF00028; cadherin; 4.				
CC	HSP: P09803; IEDH.				

KW CELL ADHESION; SIGNAL; TRANSMEMBRANE; CYTOSKELETON; GLYCOPROTEIN;
 FT CALCIUM-BINDING; REPEAT.
 FT SIGNAL 1 23 POTENTIAL.
 FT PROPE 24 49 POTENTIAL.
 FT CHAIN 50 999 DESMOGLEIN 3.
 FT DOMAIN 50 615 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 616 640 POTENTIAL.
 FT DOMAIN 641 999 CYTOSOLASMIC (POTENTIAL).
 FT REPEAT 50 158 CADHERIN 1.
 FT REPEAT 159 268 CADHERIN 2.
 FT REPEAT 269 383 CADHERIN 3.
 FT REPEAT 386 499 CADHERIN 4.
 FT REPEAT 510 935 DESMOGLEIN REPEAT 1.
 FT REPEAT 936 966 DESMOGLEIN REPEAT 2.
 FT CARBOHYD 110 110 POTENTIAL.
 FT CARBOHYD 180 180 POTENTIAL.
 FT CARBOHYD 459 459 POTENTIAL.
 FT CARBOHYD 545 545 POTENTIAL.
 SQ SEQUENCE 999 AA; 107503 MW; 4891F6AE CRC32;
 Query Match 100.0%; Score 109; DB 1; Length 999;
 Best Local Similarity 100.0%; Pred. No. 8,998-13;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 512 SARTLNRRYTPYTF 526
 QY 1 SARTLNRRYTPYTF 15
 ID IM44_YEAST STANDARD; PRT; 431 AA.
 AC 001852;
 DT 01-JUL-1993 (REL. 26, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE MITOCHONDRIAL IMPORT INNER MEMBRANE TRANSLOCASE SUBUNIT TIM44
 DE PRECURSOR (MITOCHONDRIAL PROTEIN IMPORT PROTEIN 1) (INNER MEMBRANE
 DE IMPORT SITE PROTEIN 45) (ISP45) (MEMBRANE IMPORT MACHINERY PROTEIN
 DE MIM44).
 GN TIM44 OR MP11 OR MIM44 OR ISP45 OR YII022M.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCYCETES; SACCHAROMYCETALES;
 CC SACCHAROMYCEACEAE; SACCHAROMYCES.
 PP SEQUENCE FROM N.A.
 RC STRAIN-AB320;
 RX MEDLINE: 93010953.
 RA MARSE A.C., BLOW J., GRIVELL L.A., MEIJER M.;
 RT "Mpl1, an essential gene encoding a mitochondrial membrane protein,
 RT is possibly involved in protein import into yeast mitochondria."
 RL EMO J. 11:3619-3628(1992).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
 RA CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,
 RA GENTLES S., HAMELYN N., HORSNELL T.S., HUNT S., JAGIELS K., JONES M.,
 RA LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,
 RA RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,
 RA WALSH S.V., WHITEHEAD S.;
 RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 [3]
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RX MEDLINE: 93345448.
 RA HORST M., JENSEN P., KRONIDOU N.G., BOLLIGER L., OPLIGER W.,
 RA SCHERER P., MANNING-KRUEG U., JASCUR T., SCHATZ G.;
 RT "Protein import into yeast mitochondria: the inner membrane import
 RT site protein ISP45 is the Mpl1 gene product."
 RL EMO J. 12:3035-3041(1993).
 CC -1- FUNCTION: INVOLVED IN PROTEIN IMPORT INTO THE MITOCHONDRION.
 CC AS A BINDING PROTEIN REQUIRED FOR DRIVING THE IMPORT OF
 CC PREPROTEINS. RECRUITS MITOCHONDRIAL HSP70 AND ITS CO-CHAPERONE

CC (MGE1) TO DRIVE PROTEIN TRANSLLOCATION INTO THE MATRIX USING ATP
 CC AS AN ENERGY SOURCE.
 CC -1- SUBUNIT: FORMS PART OF THE RECEPTOR COMPLEX THAT CONSISTS OF AT
 CC LEAST 3 DIFFERENT PROTEINS (TIM17, TIM23, TIM44).
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: 246881; G599977; -
 CC EMBL: X67276; G3967; -
 CC EMBL: 247047; G763324; -
 CC PIR: S25196; S25196.
 CC SGI: I0001138; MP11.
 DR SGT: I0001138; MP11.
 DR PIR: S25196; S25196.
 KM MITOCHONDRION; INNER MEMBRANE; TRANSPORT; PROTEIN TRANSPORT;
 KM TRANSLLOCATION; TRANSIT PEPTIDE; ATP-BINDING.
 FT TRANSIT 1 431 MITOCHONDRIAL IMPORT INNER MEMBRANE
 FT CHAIN 1 431
 FT NP_BIND 101 108 ATP (POTENTIAL).
 SQ SEQUENCE 431 AA; 48854 MW; 002E0771 CRC32;
 Query Match 59.6%; Score 65; DB 1; Length 431;
 Best Local Similarity 57.1%; Pred. No. 5,378-02;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 DB 13 SRTLTARYRSQYT 26
 QY 1 SARTLNRRYTPYTF 14
 ID GABR_XENLA STANDARD; PRT; 391 AA.
 AC P70005;
 DT 15-JUL-1998 (REL. 36, CREATED)
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE TRANSCRIPTION FACTOR XGATA-6B (GATA BINDING FACTOR-6B).
 GN GATA-6A OR GATA-6.
 OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
 CC MESOBATRACHIA; PIPIDAE; PIPIDAE; XENOPODINAE; XENOPUS.
 PP (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE-LIVER;
 RA GOVE C.D., MALMSLEY M., NITJAR S., BERTWISTLE D., GUILLE M.,
 RA PARTINGTON G., BOMFORD A., PATIENT R.;
 RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- FUNCTION: ASSOCIATED WITH CARDIAC SPECIFICATION AND CAN REGULATE
 CC CARDIAC-SPECIFIC TRANSCRIPTION DURING EMBRYOGENESIS. ACTIVATES
 CC THE EXPRESSION OF CARDIAC MYC-ALPHA IN VIVO.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: TO OTHER GATA-TYPE TRANSCRIPTION FACTORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Y08865; E275165; -
 CC PROSITE: PS00344; GATA_ZN_FINGER; 2.
 DR PFM: PF00320; GATA; 2.
 DR HSP: P17678; IGAD;
 KM TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING; ZINC-FINGER;
 KM NUCLEAR PROTEIN.

```

FT 2N.FING 182 206 GATA-TYPE.
FT 2N.FING 236 260 GATA-TYPE.
FT DOMAIN 70 79 POLY-SER.
FT DOMAIN 241 245 POLY-THR.
FT DOMAIN 293 298 POLY-SER.
SC SEQUENCE 391 AA; 41503 MW; 221DD428 CRC32;

Query Match
Best Local Similarity 56.9%; Score 62; DB 1; Length 391;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 120 ARPLNGSYGSPYT 132
Qy 2 ARTLNRRYTGPT 14

RESULT 4
ID HM01.CAEL STANDARD; PRT; 273 AA.
AC 018273:
LI 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE PUTATIVE HOMEBOX PROTEIN C28A5.4.
GN C28A5.4.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SEGEREMENTA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA PALMER S.;
RL SUBMITTED (APR-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -1 SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1 SIMILARITY: BELONGS TO THE DISTAL-LESS FAMILY OF HOMEBOX
CC PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
CC EMBL; Z32680; E1344298; -
CC DR WOMPEP; C28A5.4; C805325.
CC DR PROSITE; PS00027; HOMEBOX_1; 1.
CC DR PROSITE; PS50071; HOMEBOX_2; 1.
CC DR PFAM; PF00046; homeobox; 1.
CC DR HSSP; P02836; 1HDD.
CC KW HYPOTHEICAL PROTEIN; HOMEBOX; DNA-BINDING; NUCLEAR PROTEIN.
CC FT DNA_BIND 102 161 HOMEBOX.
CC ST SEQUENCE 273 AA; 30163 MW; 1655A6DF CRC32;

Query Match
Best Local Similarity 55.0%; Score 60; DB 1; Length 273;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 215 PPMTNEQYASPYLY 229
Qy 1 SARTLNRRYTGPT 15

RESULT 5
ID VG08.BPT4 STANDARD; PRT; 334 AA.
AC P19062:
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DE BASEPLATE STRUCTURAL PROTEIN GP8.
GN 8.
OS BACTERIOPHAGE T4.
OC VIRUSES; DSDNA VIRUSES; NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE;

```

CC T4-LIKE PHAGES.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-D;
RX MEDLINE; 90384864.
RA EFIMOV V.P., PRILIPOV A.G., MESYANZHINOV V.V.;
RT "Nucleotide sequences of bacteriophage T4 genes 6, 7 and 8."
RL NUCLEIC ACIDS RES. 18:5313-5313(1990).
CC -1- FUNCTION: STRUCTURAL COMPONENT OF THE BASEPLATE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X15907; G15324; -
DR PIR; JQ0658; G88PT4.
DR PIR; JF0058; JF0058.
RW STRUCTURAL PROTEIN.
SQ SEQUENCE 334 AA; 38008 MW; 8166D402 CRC32;

Query Match 55.0%; Score 60; DB 1; Length 334;
Best Local Similarity 60.0%; Pred. No. 5,92e-01;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 97 DRRYDPTF 106
OY : || ||||
6 NRRYTGPTF 15

RESULT 6 STANDARD; PIR; 304 AA.
ID C4B2.SIMV1
AC P42788;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE ZINC CARBOXYPEPTIDASE (EC 3.4.17.-) (FRAGMENT).
OS SIMULIUM VITATUM (BLACK FLY).
OC EUCARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; NEMATOCERA; CHIRONOMIDEA; SIMULIIDAE; SIMULIUM.
CN (1)
RP SEQUENCE FROM N.A.
RP TISSUE-GUT;
RX MEDLINE; 94093864.
RA RAMOS A., MAHOMAD A., JACOBS-LORENA M.;
RT "Gut-specific genes from the black fly Simulium vittatum encoding
RT trypsin-like and carboxypeptidase-like proteins."
RL INSECT MOL. BIOL. 1:149-163(1993).
CC -1- FUNCTION: INVOLVED IN THE DIGESTION OF THE BLOOD MEAL.
CC -1- TISSUE SPECIFICITY: GUT-SPECIFIC
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L08481; G161186; -
DR PROSITE; PS00132; CARBOXYPEPT_2N_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_2N_2; 1.
DR PFAM; PF00246; Zn_carboxepr; 1.
DR HSSP; P00730; ACPA.
KW HYDROLASE; CARBOXYPEPTIDASE; ZINC.
FT NON_TER 1 1
FT 58 1
FT METAL 61 61 ZINC (BY SIMILARITY).
FT 61 61 ZINC (BY SIMILARITY).

FT METAL 184 184 ZINC (BY SIMILARITY).
 FT ACT_SITE 236 236 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 259 259 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 125 148 BY SIMILARITY.
 SQ SEQUENCE 304 AA; 34849 MW; 9543CCAE CRC32;
 Query Match 54.1%; Score 59; DB 1; Length 304;
 Best Local Similarity 50.0%; Pred. No. 9.43e+01;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 Db 214 AARSAGRYGDTY 227
 :||:| 1111
 Uy 1 SAKLNNRYTGPYT 14
 RESULT 7 STANDARD; PRT: 469 AA.
 AC 027605;
 DT 15-JUL-1998 (REL. 36, CREATED)
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE NITROGENASE MOLYBDENUM-IRON PROTEIN ALPHA CHAIN (EC 1.18.6.1)
 DE NITROGENASE COMPONENT I (DINITROGENASE).
 GN NIED OR MTH1563.
 OS METHANOBACTERIUM THERMOAUTOTROPHICUM.
 OC ARCHAEA; EURARCHAEOTA; METHANOBACTERIALES; METHANOBACTERIACEAE;
 NC METHANOBACTERIUM.
 RC [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-DELTA H:
 RX MEDLINE; 98037514.
 RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,
 RA ALDREDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,
 RA HARRISON D., HOANG L., KEAGLE P., LUHM W., POTIER B., QIU D.,
 RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,
 RA JIVANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S.,
 RA MCCOUGALL S., SHIMER G., GOYAL A., PETROVSKI S., CHURCH G.M.,
 RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 strain delta: functional analysis and comparative genomics.";
 RL J. BACTERIOL. 179:7135-7155(1997).
 CC -1- FUNCTION: THE KEY ENZYMAIC REACTIONS IN NITROGEN FIXATION ARE
 CATALYZED BY THE NITROGENASE COMPLEX, WHICH HAS 2 COMPONENTS: THE
 IRON PROTEIN AND THE MOLYBDENUM-IRON PROTEIN.
 CC -1- CATALYTIC ACTIVITY: 3 REDUCED FERREDOXIN + 6 H(+) + N(2) + N ATP =
 CC 3 OXIDIZED FERREDOXIN + 2 NH(3) + N ADP + N ORTHOPHOSPHATE.
 CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS THAT BINDS
 CC 30-32 FE. 2 MO. AND INORGANIC SULFUR.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; AE000916; G2622685;
 DR PROSITE; PS00699; NITROGENASE_1.1; FALSE_NEG.
 DR PROSITE; PS00090; NITROGENASE_1.2; 1.
 DR PFAM; PF00148; Oxidored_nitro; 1.
 DR HSP; P00467; IMO.
 KW OXIDOREDUCTASE; NITROGEN FIXATION; MOLYBDENUM; IRON-SULFUR.
 SQ SEQUENCE 469 AA; 52866 MW; 08177BD4 CRC32;
 Query Match 53.2%; Score 58; DB 1; Length 469;
 Best Local Similarity 58.3%; Pred. No. 1.49e+00;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 Db 273 ARLLERYNVPY 284
 :||:| 1111
 Uy 2 ARTLNNRYTGPYT 13

RESULT 8 STANDARD; PRT: 1237 AA.
 ID POL4 DROME
 AC P10394;
 DT 01-MAR-1989 (REL. 10, CREATED)
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE RETROVIRUS-RELATED POL POLYPROTEIN FROM TRANSPOSON 412 [CONTAINS:
 DE PROTEASE (EC 3.4.22.-); REVERSE TRANSCRIPTASE (EC 2.7.7.49);
 DE ENDONUCLEASE].
 GN POL.
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYROIDEA;
 OC DROSOPHILIDAE; DROSOPHILA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 86274717.
 RA YUKI S., INOUE S., ISHIMARU S., SAIGO K.;
 RT "Nucleotide sequence characterization of a Drosophila
 RT retrotransposon, 412.";
 RL EUR. J. BIOCHEM. 158:403-410(1986).
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY 022.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; X04132; E1912;
 DR PIR; D29349; GNEF42.
 DR FLYBASE; FBgn0000006; 412.
 DR PROSITE; PS00141; ASP_PROTEASE; 1.
 DR PFAM; PF00077; IVP; 1.
 DR PFAM; PF00078; IYL; 1.
 DR HSP; P03366; IHEG.
 KW HYDROLASE; ASPARTYL PROTEASE; RNA-DIRECTED DNA POLYMERASE;
 KW ENDONUCLEASE; TRANSFERASE; POLYPROTEIN; TRANSPOSABLE ELEMENT.
 FT ACT_SITE 63 63
 SQ SEQUENCE 1237 AA; 143041 MW; 4FE4B272 CRC32;
 Query Match 53.2%; Score 58; DB 1; Length 1237;
 Best Local Similarity 42.9%; Pred. No. 1.49e+00;
 Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
 Db 1193 GHKLDFRYTGPYKT 1206
 :||:| 11111111
 Uy 2 ARTLNNRYTGPYKT 15
 RESULT 9 STANDARD; PRT: 387 AA.
 ID GAT6_CHICK
 AC P43693;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE TRANSCRIPTION FACTOR GATA-6 (GATA BINDING FACTOR-6).
 GN GAT6.
 OS GALLUS GALLUS (CHICKEN).
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
 CC ENOANTHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-WHITE LEGHORN;
 RX MEDLINE; 94365018.
 RA LAVERIERE A.C., MACNEILL C., MOELLER C., POELMANN R.E.,
 RA BURCH J.B.E., EVANS T.;
 RT "GATA-4/5/6, a subfamily of three transcription factors transcribed
 RT in developing heart and gut.";
 RL J. BIOL. CHEM. 269:23177-23184(1994).

CC -1- SUPRACELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: MORE ABUNDANT IN STOMACH, AND IN SMALL
CC INTESTINE. LOWER LEVELS IN LUNG, LIVER, OVARY AND HEART.
CC -1- SIMILARITY: TO OTHER GATA-TYPE TRANSCRIPTION FACTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U11889; G511484; -
DR PROSITE: PS00344; GATA_ZN_FINGER; 2.
DR PFAM: PF00320; GATA; 2.
DR HSSP: P17678; IGAV.
KW TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING; ZINC-FINGER;
KW NUCLEAR PROTEIN.
FT ZN_FING 181 205 GATA-TYPE.
FT ZN_FING 235 259 GATA-TYPE.
FT DOMAIN 240 244 POLY-THR.
SQ SEQUENCE 387 AA; 40249 MW; 6E67EALC CRC32;
Query Match 52.3%; Score 57; DB 1; Length 387;
Best Local Similarity 46.2%; Pred. No. 2.35e+00;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
Db 119 PRPLNGSYAPYA 131
OY 2 ARTLNRRYTGPTT 14
RESULT 10
ID CHS2_EMENTI STANDARD; PRT; 1013 AA.
AC P30584;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE CHITIN SYNTHASE 2 (EC 2.4.1.16) (CHITIN-UDP ACETYL-GLUCOSAMINYL
DE TRANSFERASE 2).
GN CHS2 OR CHSA.
OS EMERICELLA NIDULANS (ASPERGILLUS NIDULANS).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;
OC EUROTIALES; TRICHOCOMACEAE; EMERICELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PGSC 89;
RA KOUTMA N.;
RL SUBMITTED (NOV-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE OF 303-491 FROM N.A.
RX MEDLINE: 92115692.
RA BOWEN A.R., CHEN-WU J.L., WOMANT M., YOUNG R., SZANISZLO P.J.,
RA ROBBINS P.M.;
RT "Classification of fungal chitin synthases."
RT PROC. NATL. ACAD. SCI. U.S.A. 89:519-523(1992).
CC -1- FUNCTION: PLAYS A MAJOR ROLE IN CELL WALL BIOGENESIS.
CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + [1,4-(N-ACETYL-
CC BETA-D-GLUCOSAMINYL)](N) = UDP + [1,4-(N-ACETYL-BETA-D-
CC GLUCOSAMINYL)](N+1).
CC -1- SUBCELLULAR LOCATION: PLASMA MEMBRANE-BOUND.
CC -1- SIMILARITY: BELONGS TO THE CHITIN SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D21268; G465390; -

DR EMBL: M82939; G168041; -
DR PIR: C45188; C45188.
KW TRANSFERASE; GLYCOSYLTRANSFERASE; TRANSMEMBRANE; CELL WALL;
KW MULTIGENE FAMILY.
SQ SEQUENCE 1013 AA; 113678 MW; 9E03E27C CRC32;
Query Match 52.3%; Score 57; DB 1; Length 1013;
Best Local Similarity 42.9%; Pred. No. 2.35e+00;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
Db 16 ARTAESRYPERXY 29
OY 2 ARTLNRRYTGPTT 15
RESULT 11
ID IDH1_YEAST STANDARD; PRT; 360 AA.
AC P28834;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ISOCITRATE DEHYDROGENASE [NAD], MITOCHONDRIAL SUBUNIT 1 PRECURSOR
DE (EC 1.1.1.41) (ISOCITRIC DEHYDROGENASE) (NAD+-SPECIFIC IDDH).
GN IDH1 OR YN037C OR N2690.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;
OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 49-61; 72-83; 325-333 AND 339-356.
RX MEDLINE: 92355609.
RA CUPP J.R., MCALLISTER-HENN L.;
RT "Cloning and characterization of the gene encoding the IDH1 subunit
RT of NAD(+)-dependent isocitrate dehydrogenase from Saccharomyces
RT cerevisiae.";
RL J. BIOL. CHEM. 267:16417-16423(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA DUESTERHOEF A., FLOETH M., FRITZ C., HEUSS-NEITZEL D.,
RA HILBERT H., MOESTL D.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 12-27.
RC STRAIN-SG7.
RX MEDLINE: 90330530.
RA KEYS D.A., MCALLISTER-HENN L.;
RT "Subunit structure, expression, and function of NAD(H)-specific
RT isocitrate dehydrogenase in Saccharomyces cerevisiae.";
RL J. BACTERIOL. 172:4280-4287(1990).
RN [4]
RP RNA-BINDING.
RX MEDLINE: 94089379.
RA ELZINGA S.D.J., BEDNARZ A.L., VAN OOSTERUM K., DEKKER P.J.T.,
RA GRIEVEL L.A.;
RT "Yeast mitochondrial NAD(+)-dependent isocitrate dehydrogenase is an
RT RNA-binding protein.";
RL NUCLEIC ACIDS RES. 21:5328-5331(1993).
CC -1- FUNCTION: PERFORMS AN ESSENTIAL ROLE IN THE OXIDATIVE FUNCTION OF
CC THE CITRIC ACID CYCLE. ALSO BINDS RNA; SPECIFICALLY TO THE
CC 5'-UNTRANSLATED LEADERS OF MITOCHONDRIAL MNAS.
CC -1- CATALYTIC ACTIVITY: ISOCITRATE + NAD(+) = 2-OXOGLOUTARATE +
CC CO(2) + NADH.
CC -1- ENZYME REGULATION: ALLOSTERICALLY REGULATED BY SEVERAL COMPOUNDS
CC INCLUDING AMP, NAD+, AND CITRATE.
CC -1- SUBUNIT: OCTAMER OF TWO NONIDENTICAL SUBUNITS IDH1 AND IDH2.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- SIMILARITY: BELONGS TO THE ISOCITRATE AND ISOPROPYLMALATE
CC DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M95203; G171766; -

DR EMBL: 271313; E239881; -

DR PIR: S31264; S31264.

DR SGD: L0000849; IDH1.

DR PROSITE: PS00470; IDH_IMDH; 1.

DR PFAM: PF00180; Isocdh; 1.

DR HSSP: P00351; 10SI.

DR OXIDOREDUCTASE; NAD; TRICARBOXYLIC ACID CYCLE; TRANSIT PEPTIDE;

KW MITOCHONDRION; ALLOSTERIC ENZYME; RNA-BINDING.

FT TRANSIT 1 11

FT CHAIN 12 360

FT ACT_SITE 103 103

FT BINDING TO ISOCITRATE SUBUNIT 1.

FT BINDING TO ISOCITRATE (BY SIMILARITY).

SO SEQUENCE 360 AA; 39324 MW; 706BCA76 CRC32;

Query Match 51.4%; Score 56; DB 1; Length 360;

Best Local Similarity 42.9%; Pred. No. 3,68e+00;

Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 17 AERTLPKYGGRFT 30

QY 1 SARTLNRRITGPT 14

RESULT 12

ID N145 YEAST STANDARD; PRT; 1317 AA.

AC P49687;

DT 01-FEB-1996 (REL. 33, CREATED)

DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE NUCLEOPORIN NUP145 (NUCLEAR PORE PROTEIN NUP145).

GN NUP145 OR YGL092W.

OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).

OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCETES; SACCAROMYCETALES;

OC SACCHAROMYCETACEAE; SACCAROMYCES.

RN [1]

RN SEQUENCE FROM N.A.

RX MEDLINE: 94320139.

RA "Nup145p is required for nuclear export of mRNA and binds homopolymeric RNA in vitro via a novel conserved motif.";

RT Homopolymeric RNA in vitro via a novel conserved motif.";

RL CELL 78:275-289(1994).

RN [2]

RN SEQUENCE FROM N.A.

RC STRAIN-ATCC 26109 / X2180;

RX MEDLINE: 94253246.

RA WENTE S.R.; BLOBEL G.;

RT "NUP145 encodes a novel yeast glycine-leucine-phenylalanine-glycine (GLFG) nucleoporin required for nuclear envelope structure.";

RL J. CELL BIOL. 125:955-969(1994).

RN [3]

RN SEQUENCE FROM N.A.

RX STRAIN-S288C;

RA MEDLINE: 97435481.

RA RIEGER M.; BRUECKNER M.; SCHAEFER M.; MUELLER-AUER S.;

RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae chromosome VII.";

RT Chromosome VII.";

RL YEAST 13:1077-1090(1997).

CC -1- FUNCTION: MAY PLAY A DIRECT ROLE IN NUCLEOCYTOPLASMIC EXPORT OF RNA AND ALSO IN PROTEIN IMPORT. BINDS HOMOPOLYMERIC RNA.

CC -1- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX.

CC -1- DOMAIN: CONTAINS G-L-F-G REPEATS.

CC -1- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X76557; G505271; -

DR EMBL: 232672; G496731; -

DR EMBL: 272614; E243975; -

DR SGD: L0001294; NUP145.

KW NUCLEAR PROTEIN; TRANSPORT; REPEAT; RNA-BINDING.

FT CONFLICT 281 282

FT CONFLICT 1142 1142

FT CONFLICT 1310 1317

FT CONFLICT 145660 MW; F4CA0DB4 CRC32;

SO SEQUENCE 1317 AA; 145660 MW; F4CA0DB4 CRC32;

Query Match 51.4%; Score 56; DB 1; Length 1317;

Best Local Similarity 46.2%; Pred. No. 3,68e+00;

Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 1145 NQALKREYGNV 1157

QY 1 SARTLNRRITGPT 13

RESULT 13

ID GVPLANAFI STANDARD; PRT; 245 AA.

AC P55149;

DT 01-OCT-1996 (REL. 34, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE GVPE/L PROTEIN.

GN GVPE/L.

OS ANABAENA FLOS-AQUAE.

OC BACTERIA; CYANOBACTERIA; NOSTOCALES; NOSTOCACEAE; ANABAENA.

RN [1]

RN SEQUENCE FROM N.A.

RC STRAIN-CCAP 1403/13F;

RX MEDLINE: 97217685.

RA KINSMAN R.; HAYES P.K.;

RT "Genes encoding proteins homologous to halobacterial Gyps N, J, K, F & L are located downstream of gypC in the cyanobacterium Anabaena flos-aquae.";

RL DNA SEQ. 7:97-106(1997).

CC -1- FUNCTION: MAY PLAY A STRUCTURAL OR REGULATORY ROLE IN GAS VESICLE SYNTHESIS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL: U17109; G595974; -

KW GAS VESICLE.

SO SEQUENCE 245 AA; 28676 MW; F46BD20E CRC32;

Query Match 50.5%; Score 55; DB 1; Length 245;

Best Local Similarity 53.8%; Pred. No. 5,72e+00;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 228 RIRYNNFTAYTF 240

QY 3 RILNRRITGPT 15

RESULT 14

ID Y4ZC RHISN STANDARD; PRT; 261 AA.

AC P55730;

DT 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE HYPOTHETICAL 28.3 KD PROTEIN Y4ZC.

GN Y4ZC.

OS RHIZOBIUM SP. (STRAIN NGR234).

OG PLASMID SYN PNC8234A.
 OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
 OC RHIZOBIACEAE; RHIZOBIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 97305956.
 RA FREIERG C.A., FELLAY R., BAIRROCH A., BROUGHTON W.J., ROSENTHAL A.,
 RA PERRET X.;
 RT "Molecular basis of symbiosis between Rhizobium and legumes.";
 RL NATURE 387:394-401(1997).
 CC -1- SIMILARITY: TO P.SYRINGAE (PV. PHASEOLICOLA) AVIRULENCE PROTEIN
 CC AVIRPH3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000109; G2182749; -
 CC KX HYPOPHETICAL PROTEIN: PLASMID.
 SQ SEQUENCE 261 AA; 28349 MW; 51A967AE CRC32;
 DB 238 RSLANR5NP 247
 OY 3 RLNNRYTGP 12
 Query Match 50.5%; Score 55; DB 1; Length 261;
 Best Local Similarity 60.0%; Pred. No. 5.72e+00;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 RESULT 15
 ID G6A.XENIA STANDARD; PRT; 391 AA.
 AC Q91678;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE TRANSCRIPTION FACTOR XGATA-6A (GATA BINDING FACTOR-6A).
 GN GATA-6A OR GATA-6.
 OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
 OC EKAROTIA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
 OC MESOBATRACHIA; PIPIDAE; PIPIDAE; XENOPODINAE; XENOPUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE: HEART, AND GUT;
 RX MEDLINE: 96175597.
 RA JIANG Y., EVANS T.;
 RT "The xenopus GATA-4/5/6 genes are associated with cardiac
 RT specification and can regulate cardiac-specific transcription during
 RT embryogenesis.";
 RL DEV. BIOL. 174:258-270(1996).
 CC -1- FUNCTION: ASSOCIATED WITH CARDIAC SPECIFICATION AND CAN REGULATE
 CC CARDIAC-SPECIFIC TRANSCRIPTION DURING EMBRYOGENESIS. ACTIVATES
 CC THE EXPRESSION OF CARDIAC MHC-ALPHA IN VIVO.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN HEART, SMALL
 CC INTESTINE, AND STOMACH. FOUND AT LOWER LEVELS IN LUNG, PANCREAS,
 CC AND COLON.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN CARDIAC PROGENITORS DURING
 CC EMBRYOGENESIS AND UPREGULATED DURING GASTRULATION.
 CC -1- SIMILARITY: TO OTHER GATA-TYPE TRANSCRIPTION FACTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: U45454; G1209880; -
 DR PROSITE: PS00344; GATA_ZN_FINGER; 2.
 DR PFAM: PF00320; GATA; 2.
 DR HSSP: P1678; 1GAD.
 KW TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING; ZINC-FINGER;
 KW NUCLEAR PROTEIN.
 FT ZN_FING 182 206 GATA-TYPE.
 FT FT 260 260 GATA-TYPE.
 FT DOMAIN 293 298 POLY-SER.
 SQ SEQUENCE 391 AA; 41780 MW; 1BC6C528 CRC32;
 DB 120 ARSLNGSGSHY 132
 OY 2 ARTLNRYTGP 14
 Query Match 50.5%; Score 55; DB 1; Length 391;
 Best Local Similarity 53.8%; Pred. No. 5.72e+00;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 Search completed: Fri Jun 11 17:45:23 1999
 Job time : 8 secs.

RX MEDLINE: 94107601.
 RA MURPHY E., KOSBER B.T., GEORGES-CORREBOT M.C., YON B., PINTER A.,
 RA COOK D., KIENY M.P., GEORGES A., MATHIOT C., BARRE-SINOSSI F.,
 RA GIRARD M.,
 RT "Diversity of V3 region sequences of human immunodeficiency viruses
 type 1 from the central African Republic";
 RL AIDS RES. HUM. RETROVIRUSES 9:997-1006(1993).
 DR EMBL: L11498; G305527; .
 KW ENVELOPE PROTEIN.
 FT NON_TER
 SQ SEQUENCE 80 AA; 9239 MW; 15E98296 CRC32;

Query Match 52.3%; Score 57; DB 14; Length 80;
 Best Local Similarity 70.0%; Pred. No. 7.42e+00;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 63 NTRYTGSYTL 72
 | |||||:
 QY 6 NNTYTGPT 15

RESULT 15
 ID 005776 PRELIMINARY; PRT; 175 AA.
 AC 005776;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE SIMILAR TO MSF1 PROTEIN.
 GN L8167.12.
 OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCAROMYCETALES;
 OC SACCAROMYCETACEAE; SACCAROMYCETES.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C (AB972);
 RX MEDLINE: 97313267.
 RA JOHNSTON M., HILLIER L., RILES L., ALBERMANN K., ANDRE B., ANSGORGE W.,
 RA BENES V., BRUCKNER M., DELTUS H., DUBOIS E., DUSTERHOFT A.,
 RA EUTIAN K.D., FLOETH M., GOFFEAU A., HEBLING U., HEUMANN K.,
 RA HEUSS-NEITZEL D., HILBERT H., HILGER F., KLEINE K., KOTTER P.,
 RA LOUIS E.J., MESSENGUY F., MEWES H.W., MOSGA T., MOSTL D.,
 RA MILLER-AUER S., NENTWICH U., OBERMAYER B., PIRAVANDI E., POHL T.M.,
 RA PORTERELLE D., PURTELLE B., RECHMANN S., RIEGER M., RINKE M., ROSE M.,
 RA SCHARPE M., SCHERENS B., SCHOLLER P., SCHWAGER C., SCHWARZ S.,
 RA UNDERWOOD A.P., URRESTARAZU L.A., VANDENBOL M., VERHASSELT P.,
 RA VIERENDELS F., VOET M., VOLCKAERT G., VOSS H., WAMBUYT R., WEDLER E.,
 RA WEDLER H., ZIMMERMANN F.K., ZOLLNER A., HANI J., HOEHLISSEL J.D.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
 XII";
 RL NATURE 387:0-0(0).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C (AB972);
 RA PAULEY A.;
 RL SUBMITTED (SEP-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C (AB972);
 RA WATERSTON R.;
 RL SUBMITTED (SEP-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN (4)
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C (AB972);
 RA CHERRY J.M.;
 RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: U14913; G544509; .
 SQ SEQUENCE 175 AA; 20108 MW; 9FCA0042 CRC32;

Query Match 52.3%; Score 57; DB 3; Length 175;
 Best Local Similarity 46.2%; Pred. No. 7.42e+00;
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 19 SRAFFNRYPNPYS 31

QY 2 ARTLNNTYTGPT 14

Search completed: Fri Jun 11 17:42:47 1999
 Job time : 123 secs.

(TM)

Distribution rights by Oxford Molecular Ltd

protein database search, using Smith-Waterman algorithm

Fri Jun 11 17:52:50 1999; MasPar time 4.77 Seconds

output not generated.

1 QSGTMRHSTGGTN 15

PAM 150

170751 seqs, 21266608 residues

Listing first 45 summaries

a-geneseq35

39:part39

mean 17.125; Variance 54.917; scale 0.312

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No
1	99	100.0	15	20	W04847	Self epitope of desmo	5.56e-04
2	99	100.0	6	6	R30742	Human pemphigus vulva	5.56e-04
3	54	54.5	172	1	R06023	Viral haemagglutinin	5.66e+01
4	49	49.5	440	2	R10423	Nod L gene product.	1.82e+02
5	48	48.5	802	11	R56550	Cold acclimatization	1.52e+02
6	48	48.5	830	37	W80314	CIIRA (class II) trans	2.29e-02
7	48	48.5	987	31	W57057	Class II trans activa	2.29e-02
8	48	48.5	1085	3	P61046	CIIRA (class II) trans	2.29e-02
9	48	48.5	1106	37	W8312	A. nidulans phosphoenol	2.29e-02
10	48	48.5	1130	31	W57056	Class II trans activa	2.29e-02
11	48	48.5	1130	37	W80313	CIIRA (class II) trans	2.29e-02
12	48	48.5	1130	14	R72452	Class II transactivat	2.29e-02
13	48	48.5	1130	16	R81569	Class II transactivat	2.29e-02
14	48	48.5	1307	37	W80311	CIIRA (class II) trans	2.29e-02
15	47	47.5	483	38	W73510	Alpha amylase protein	2.87e+02
16	47	47.5	483	4	R2434	Alpha amylase mutant	2.87e+02

17	47	47.5	483	11	R58743	W138F alpha-amylase.	2.87e+02
18	47	47.5	483	11	R58012	M197Y alpha-amylase.	2.87e+00
19	47	47.5	483	11	R58731	M15P alpha-amylase.	2.87e+00
20	47	47.5	483	11	R58011	M197F alpha-amylase.	2.87e+02
21	47	47.5	483	22	W05232	Bacillus licheniform	2.87e+02
22	47	47.5	483	11	R57986	M197F alpha-amylase.	2.87e+00
23	47	47.5	483	11	R58745	W138A alpha-amylase.	2.87e+00
24	47	47.5	483	11	R58741	M15W alpha-amylase.	2.87e+00
25	47	47.5	483	22	W08203	B. licheniformis mutan	2.87e+02
26	47	47.5	483	22	W08205	B. licheniformis mutan	2.87e+02
27	47	47.5	483	33	W57972	Alpha-amylase mutant	2.87e+02
28	47	47.5	483	33	W57973	Alpha-amylase mutant	2.87e+02
29	47	47.5	483	36	W80194	Bacillus licheniformi	2.87e+02
30	47	47.5	483	22	W05230	Bacillus licheniformi	2.87e+02
31	47	47.5	483	22	W05228	Bacillus licheniformi	2.87e+02
32	47	47.5	483	11	R58747	W138H alpha-amylase.	2.87e+02
33	47	47.5	483	11	R58748	W138C alpha-amylase.	2.87e+02
34	47	47.5	483	11	R58727	M15N alpha-amylase.	2.87e+02
35	47	47.5	483	2	R10576	Mutant alpha-amylase	2.87e+02
36	47	47.5	483	11	R58736	M15O alpha-amylase.	2.87e+02
37	47	47.5	487	36	W80189	A4 form of the alpha-	2.87e+02
38	47	47.5	487	36	W80215	A4 form alpha-amylase	2.87e+02
39	47	47.5	487	36	W80193	A4 form alpha-amylase	2.87e+02
40	47	47.5	487	32	W05234	Bacillus licheniformi	2.87e+02
41	47	47.5	487	22	W05236	Bacillus licheniformi	2.87e+02
42	47	47.5	512	38	W13509	Alpha-amylase protein	2.87e+02
43	47	47.5	560	25	W22522	Alpha-amylase-cellul	2.87e+02
44	47	47.5	630	25	W22523	Alpha-amylase-CenA ce	2.87e+02
45	47	47.5	630	25	W23603	Alpha-amylase-CenA ce	2.87e+02

ALIGNMENTS

Query Match 100.0%: Score 99; DB 20; Length 15;

Best Local Similarity 100.0%; Pred. No. 5.56e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 qsgtmrttrhstgctn 15
|||||
QY 1 QSGTMRTRHSTGCTN 15

RESULT 2
ID R30742 standard; Protein; 999 AA.
AC R30742;
DT 14-JUN-1993 (first entry)
DE Human pemphigus vulgaris 130KD antigen.
KW Pemphigus vulgaris; skin disease; autoantibodies;
KM keratinocyte cell surface antigen; glycoprotein; cell adhesion.
OS Homo sapiens.
PN US7798918-A.
PD 15-DEC-1992.
PF 27-NOV-1991; 798918.
PR 27-NOV-1991; US-798918.
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
PI Amagai M, Klaus-Kovtun V, Stanley JR;
DR WPI; 93-067436/08.
DR N-PSDB; Q35992.
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for
diagnostic and therapeutic uses
PS Disclosure; Fig 7; 50pp; English.
CC This sequence is the pemphigus vulgaris 130KD antigen. The protein
and its encoding DNA may be used in the diagnosis and treatment of
pemphigus vulgaris. It is thought that the antigen may be a cell
adhesion molecule.
CC Sequence 999 AA;

Query Match 100.0%; Score 99; DB 6; Length 999;
Best Local Similarity 100.0%; Pred. No. 5.56e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 762 qsgtmrttrhstgctn 776
|||||
QY 1 QSGTMRTRHSTGCTN 15

RESULT 3
ID R06023 standard; Protein; 572 AA.
AC R06023;
DT 04-DEC-1990 (first entry)
DE Viral haemagglutinin neuraminidase.
KW Haemagglutinin neuraminidase; vaccine; parainfluenza; ds.
OS Bovine parainfluenza type III virus.
PN J02156883-A.
PD 15-JUN-1990.
PF 09-DEC-1988; 311655.
PR 09-DEC-1988; JP-311655.
PA (JAPG) NIPPON ZEON KK.
DR WPI; 90-228484/30.
DR N-PSDB; Q05319.
PT Recombinant vaccinia virus - is which all or part of DNA coding
PT membrane fusion protein in combined to genom region.
PS Disclosure; P; Japanese.
CC Fragment of parainfluenza viral membrane fusion protein encoding DNA
CC may be incorporated into the vaccina virus, which may then be used
CC as a live vaccine for cows.
SO Sequence 572 AA;

Query Match 54.5%; Score 54; DB 1; Length 572;
Best Local Similarity 53.3%; Pred. No. 5.66e+01;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 16 epgttrhskatn 30
::|||
QY 1 QSGTMRTRHSTGCTN 15

RESULT 4
ID R10423 standard; Protein; 140 AA.

AC R10423;
DT 09-APR-1991 (first entry)
DE Nod L gene product.
KW Nod genes; parasite; toxin; promoter; nodulation; flavones; nod L.
OS Bradyrhizobium japonicum.
PN US4963519-A.
PD 08-JAN-1991.
PF 24-JUL-1987; 077561.
PR 24-JUL-1987; US-077561.
PA (STAC) STACEY G.
PI Stacey G, Schell MG, Nieuwkoop AJ, Deshmane NA, Banfaiyi Z;
DR WPI; 91-036325/05.
DR N-PSDB; Q10325.
PT Nod genes and control sequences of Bradyrhizobium japonicum -
used for improving nodulation and for prodn. of foreign protein,
e.g. plant parasite toxin
PS Disclosure; Page 4; 11pp; English.
CC The nod L gene product may be used to restore nodulating ability
CC to Nod- mutants of Bradyrhizobium or Rhizobium strains and can
CC enhance nodulating ability.
SO Sequence 140 AA;

Query Match 49.5%; Score 49; DB 2; Length 140;
Best Local Similarity 50.0%; Pred. No. 1.82e+02;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 85 skrrhatagtk 96
::|||
QY 4 TWRTTRHSTGCTN 15

RESULT 5
ID R56550 standard; Protein; 802 AA.
AC R56550;
DT 20-FEB-1995 (first entry)
DE Cold acclimatization protein CAP160 from spinach leaf.
KW Cold acclimatization; cold-tolerance; transgenic plant; drought-;
KM drought-resistance; drought-tolerance; crop improvement; CAP160;
KW freezing-tolerance.
OS Spinacia oleracea.
PN W09417186-A.
PD 04-AUG-1994.
PF 21-JAN-1994; U00581.
PR 21-JAN-1993; US-007107.
PA (UYFL) UNIV FLORIDA.
PI Guy CL, Haskell DW, Hoffig A, Neven LG;
DR WPI; 94-264100/32.
DR N-PSDB; Q70904.
PT Cold acclimatization proteins CAP85 and CAP160 and their
PT nucleotide sequences - used to confer increased cold tolerance
PT and drought resistance on plants and microorganisms by genetic
PT transformation
PS Claim 6; Page 29; 49pp; English.
CC Transgenic plants expressing this cold acclimatization
CC protein have improved cold tolerance and/or drought-resistance.
CC The cell expressing the protein is from the family Solanaceae,
CC a citrus plant, a bacterium or a yeast cell.
SO Sequence 802 AA;

Query Match 48.5%; Score 48; DB 1; Length 802;
Best Local Similarity 53.3%; Pred. No. 2.39e+02;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 768 qdtgtrhltgtg 782
|||||
QY 1 QSGTMRTRHSTGCTN 15

RESULT 6
ID W80314 standard; Protein; 830 AA.
AC W80314;

DE 01-FEB-1999 (first entry)
 DE CIIITA (class II transactivator) type IV protein.
 KW CIIITA gene; class II transactivator; MHC class II molecule;
 KW Interferon-gamma; Interleukin-4; vaccine; cancer treatment.
 PN Homo sapiens.
 EP-874049-A1.
 PD 28-OCT-1998.
 PR 21-APR-1998; 400968.
 PR 22-APR-1997; FR-004954.
 PA (TRGE) TRANSGENE SA.
 PI Mach B;
 DR WPI; 98-55915/48.
 PT DNA encoding MHC class II trans-activator polypeptides) - useful
 PI e.g. as primers for enzymatic amplification, as detection probes or
 PT as inhibitors of expression of genes encoding MHC in dendritic cells
 PS Disclosure; pages 66-69; 86pp; French.
 CC The present sequence represents type IV CIIITA (class II transactivator)
 CC protein. The products can be used to treat diseases for which
 CC enhanced expression of genes coding for MHC class II molecules is
 CC desired, especially where the enhanced expression is desired in
 CC dendritic cells or after induction by a cytokine, especially
 CC Interferon-gamma or Interleukin-4. Inhibitors of the gene or protein
 CC can be used to treat diseases for which reduced expression of genes
 CC coding for MHC class II molecules is desired or can be used as vaccines,
 CC especially for cancer treatment.
 SQ Sequence 830 AA;

Query Match 48.5%; Score 48; DB 37; Length 830;
 Best Local Similarity 55.6%; Pred. No. 2.29e+02;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 497 gpgllrarq 505
 :||:|:|:
 QY 1 QSGTMRTRH 9

RESULT 7
 ID W57057 standard; Protein; 987 AA.
 AC W57057;
 DE 17-AUG-1998 (first entry)
 DE Class II trans activator (CIIITA) 151 deletion mutant.
 KW Class II trans activator; MHC; CIIITA; autoimmune disease; treatment;
 KW transplantation; xenograft; major histocompatibility complex; diabetes;
 KW gene therapy; arthritis; ribozyme; 151 deletion mutant.
 OS Homo sapiens.
 OS Synthetic.
 FT Key Location/Qualifiers
 FT Misc_difference 6 /note- "wild type 143 amino acid residues deletion
 after this position"
 FT Misc_difference 7 /label- E150I
 FT /note- "wild type Glu is replaced by Ile"
 FT Misc_difference 8 /label- L151I
 FT /note- "wild type Leu is replaced by Ile"
 PN W09815626-A2.
 PD 16-APR-1998.
 PR 08-OCT-1997; G02751.
 PR 21-MAR-1997; GB-005911.
 PR 08-OCT-1996; GB-020940.
 PA (CHIL-) INST CHILD HEALTH.
 PI Fabre JM, Gustafsson KT, Yun S;
 DR N-PSDB: V28717.
 PT Deleted form of the class II trans-activator that reduces class II
 PT antigen expression - and ribozyme directed to trans-activator mRNA,
 PT related nucleic acid, vectors, transformed cells and antibodies,
 PT used for treating auto-immune disease and to inhibit xeno-graft
 rejection
 PS Claim 15: Pages 69-72; 104pp; English.
 CC This represents a 151 deletion mutant of the class II trans activator
 CC (CIIITA) polypeptide. The N-terminal region of the wild type CIIITA is

CC deleted and can be used for reducing the expression of major
 CC histocompatibility complex (MHC) class II antigens. A ribozyme targeted
 CC to bases 1159-1161 (GDU) of human CIIITA mRNA, or corresponding target in
 CC other species and the CIIITA polypeptide can be used to reduce expression
 CC of MHC class II antigens, particularly for treatment of autoimmune
 CC disease (e.g. arthritis and diabetes) or to treat non-human animals
 CC intended as source of xenografts. The ribozyme and the CIIITA polypeptide
 CC may be generated in vivo by gene therapy, using the encoding nucleic
 CC acids targeted for localised suppression of the immune response. Material
 CC from transgenic animals in which at least some cells are stably
 CC transfected with the CIIITA encoding nucleic acids are used for animal-to-
 CC human transplantation. The CIIITA polypeptide suppresses synthesis of
 CC class II antigens, in cells that express them constitutively or after
 CC lymphokine induction. It has no transcription activity but still binds to
 CC regulatory proteins in the promoter region of class II genes, so
 CC competitively inhibits endogenous CIIITA.
 SQ Sequence 987 AA;

Query Match 48.5%; Score 48; DB 31; Length 987;
 Best Local Similarity 55.6%; Pred. No. 2.29e+02;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 654 gpgllrarq 662
 :||:|:|:
 QY 1 QSGTMRTRH 9

RESULT 8
 ID P61048 standard; Protein; 1085 AA.
 AC P61048;
 DE 03-SEP-1991 (first entry)
 DE A. nidulans phosphoenol pyruvic acid carboxylase gene product.
 KW PPC; photosynthesis; PA181; PA171; PA172; PA172A.
 OS Anacystis nidulans.
 PN J61056082-A.
 PD 20-MAR-1986.
 PR 24-AUG-1984; 177352.
 PR 24-AUG-1984; JP-177352.
 PA (KOHZ/) KOHZUKI H.
 DR WPI; 86-115948/18.
 DR N-PSDB: N60838.
 PT PPC gene of cloned photosynthetic microorganisms - for
 PT intensifying photosynthesis carbon dioxide-fixing ability of
 PT bacteria.
 PS Disclosure; Fig 1; 16pp; Japanese.
 CC Transforming a bacteria with the PPC encoding sequence intensifies
 CC CO2 fixation ability. Other PPC genes may also be found by
 CC hybridisation with an A. nidulans derived probe.
 SQ Sequence 1085 AA;

Query Match 48.5%; Score 48; DB 3; Length 1085;
 Best Local Similarity 54.5%; Pred. No. 2.29e+02;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 1054 tglmrtryxk 1064
 :||:|:|:
 QY 2 SGTMRTRHSTG 12

RESULT 9
 ID W80312 standard; Protein; 1106 AA.
 AC W80312;
 DE 01-FEB-1999 (first entry)
 DE CIIITA (class II transactivator) type II protein.
 KW CIIITA gene; class II transactivator; MHC class II molecule;
 KW Interferon-gamma; Interleukin-4; vaccine; cancer treatment.
 OS Homo sapiens.
 EP-874049-A1.
 PD 28-OCT-1998.
 PR 21-APR-1998; 400968.
 PR 22-APR-1997; FR-004954.
 PA (TRGE) TRANSGENE SA.
 PI Mach B;

DR MPI: 98-559115/48.
PT DNA encoding MHC class II trans-activator polypeptide(s) - useful
PT e.g. as primers for enzymatic amplification, as detection probes or
PT as inhibitors of expression of genes encoding MHC in dendritic cells
PS Claim 10: Pages 58-61; 86pp; French.
CC The present sequence represents type II CIITA (class II transactivator)
CC protein. The products can be used to treat diseases for which
CC enhanced expression of genes coding for MHC class II molecules is
CC desired, especially where the enhanced expression is desired in
CC dendritic cells or after induction by a cytokine, especially
CC interferon-gamma or interleukin-4. Inhibitors of the gene or protein
CC can be used to treat diseases for which reduced expression of genes
CC coding for MHC class II molecules is desired or can be used as vaccines,
CC especially for cancer treatment.
SQ Sequence 1106 AA;

Query Match 48.5%; Score 48; DB 37; Length 1106;
Best Local Similarity 55.6%; Pred. No. 2.29e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 773 gpgtlrarg 781
1:1:1:1:1:
1 OSGTMRTRH 9

RESULT 10
ID W57056 standard; Protein; 1130 AA.
AC W57056;
DT 17-AUG-1998 (first entry)
DE Class II trans activator; MHC; CIITA; autoimmune disease; treatment;
KW Class II trans activator; MHC; CIITA; autoimmune disease; diabetes;
KW transplantation; xenograft; major histocompatibility complex; diabetes;
KW gene therapy; arthritis; ribozyme.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Region 52..137
FT Region /note="N-terminal region rich in glutamate/aspartate"
FT Region 163..195
FT Region /note="region rich in proline/serine/threonine"
FT Region 209..237
FT Region /note="region rich in proline/serine/threonine"
FT Region 261..322
FT Region /note="region rich in proline/serine/threonine"
PN MO9815626-A2.
PD 16-APR-1998.
PF 08-OCT-1997; G02751.
PR 21-MAR-1997; GB-005911.
PR 08-OCT-1996; GB-020940.
PA (CHIL-) INST CHILD HEALTH.
PI Fabre JM, Gustafsson KT, Yun S;
DR MPI: 98-240813/21.
DR N-PSDB; V28716.
PT Deleted form of the class II trans-activator that reduces class II
PT antigen expression - and ribozyme directed to trans-activator mRNA,
PT related nucleic acid, vectors, transformed cells and antibodies,
PT used for treating auto-immune disease and to inhibit xeno-graft
PT rejection
PS Claim 14; Fig 5; 104pp; English.
CC This represents a class II trans activator (CIITA) polypeptide. When
CC the N-terminal region of the CIITA is deleted, it can be used for
CC reducing the expression of major histocompatibility complex (MHC) class
CC II antigens. A ribozyme targeted to bases 1159-1161 (GUA) of human CIITA
CC mRNA, or corresponding target in other species and the CIITA polypeptide
CC can be used to reduce expression of MHC class II antigens, particularly
CC for treatment of autoimmune disease (e.g. arthritis and diabetes) or to
CC treat non-human animals intended as source of xenografts. The ribozyme
CC and the CIITA polypeptide may be generated in vivo by gene therapy, using
CC the encoding nucleic acids targeted for localised suppression of the
CC immune response. Material from transgenic animals in which at least some
CC cells are stably transfected with the CIITA encoding nucleic acids are
CC used for animal-to-human transplantation. The CIITA polypeptide
CC suppresses synthesis of class II antigens. In cells that express them
CC constitutively or after lymphokine induction. It has no transcription

CC activity but still binds to regulatory proteins in the promoter region of
CC class II genes, so competitively inhibits endogenous CIITA.
SQ Sequence 1130 AA;

Query Match 48.5%; Score 48; DB 31; Length 1130;
Best Local Similarity 55.6%; Pred. No. 2.29e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 797 gpgtlrarg 805
1:1:1:1:1:
1 OSGTMRTRH 9

RESULT 11
ID W80313 standard; Protein; 1130 AA.
AC W80313;
DT 01-FEB-1999 (first entry)
DE CIITA (class II transactivator) type III protein.
KW CIITA gene; class II transactivator; MHC class II molecule;
KW Interferon-gamma; Interleukin-4; vaccine; cancer treatment.
OS Homo sapiens.
PN EP-874049-A1.
PD 28-OCT-1998.
PF 21-APR-1998; 400968.
PR 22-APR-1997; FR-004954.
PA (TRGE) TRANSENE SA.
PI Mach B;
DR MPI: 98-559115/48.
PT DNA encoding MHC class II trans-activator polypeptide(s) - useful
PT e.g. as primers for enzymatic amplification, as detection probes or
PT as inhibitors of expression of genes encoding MHC in dendritic cells
PS Claim 31; Pages 62-66; 86pp; French.
CC The present sequence represents type III CIITA (class II transactivator)
CC protein. The products can be used to treat diseases for which
CC enhanced expression of genes coding for MHC class II molecules is
CC desired, especially where the enhanced expression is desired in
CC dendritic cells or after induction by a cytokine, especially
CC interferon-gamma or interleukin-4. Inhibitors of the gene or protein
CC can be used to treat diseases for which reduced expression of genes
CC coding for MHC class II molecules is desired or can be used as vaccines,
CC especially for cancer treatment.
SQ Sequence 1130 AA;

Query Match 48.5%; Score 48; DB 37; Length 1130;
Best Local Similarity 55.6%; Pred. No. 2.29e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 797 gpgtlrarg 805
1:1:1:1:1:
1 OSGTMRTRH 9

RESULT 12
ID R72452 standard; Protein; 1130 AA.
AC R72452;
DT 08-DEC-1995 (first entry)
DE Class II transactivator protein.
KW Class II transactivator protein; MHC class II;
KW Insulin dependent diabetes; multiple sclerosis; lupus erythematosus;
KW Rheumatoid arthritis; Kozak box; ATP/GTP binding cassette;
KW yeast RNA binding protein; transcription activation domain.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Region 52..137
FT Region /note="acidic region rich in Glu and Asp"
FT Region 163..195
FT Region /note="region I rich in Pro/Ser/Thr"
FT Region 209..237
FT Region /note="region II rich in Pro/Ser/Thr"
FT Region 261..322
FT Region /note="region III rich in Pro/Ser/Thr"
FT binding_site 420..426
FT /label= ATP/GTP binding site

PD EP-648835-A.
 PN 19-APR-1995.
 PF 26-AUG-1994; 113378.
 PR 26-AUG-1993; EP-113655.
 PA (MACH/) MACH B F.
 PI MACH BF;
 DR N-PSDB; 088110.
 PT Gene encoding a protein displaying class II trans-activator
 PT activity - useful in compsn. for treating diseases associated
 PS with impaired expression of MHC-II genes, etc.
 PS Claim 11: Page 17-21; 32pp; English.
 CC This protein sequence is encoded by the gene represented in 088110.
 CC The protein is a class II transactivator (CIITA) essential for the
 CC control of MHC class II gene expression in B lymphocytes. The DNA
 CC sequence has two in frame start codons. The first of these is at position
 CC 116 and acts as the translation initiation site. The second is present
 CC (at position 188) in the context of a perfect Kozak box and may also
 CC serve as a start codon for a protein of 1106 amino acids. The protein
 CC contains three stretches rich in proline/serine/threonine. It also
 CC contains a region rich in glutamate/aspartate (an acidic region) and an
 CC AIP/GIP binding cassette. The acidic regions followed by three stretches
 CC rich in proline, serine and threonine resembles a transcription
 CC activation domain. There is also a leucine-rich region around amino
 CC acids 979 to 1061 that shows weak homology with the N terminal portion of
 CC a yeast RNA binding protein. The protein can be used for the
 CC identification of inhibitors to proteins with CIITA activity. The
 CC sequence can be used to treat diseases where a decrease in the level of
 CC expression of MHC cells II genes is desired e.g. insulin dependant
 CC diabetes, multiple sclerosis, lupus erythematosus and rheumatoid
 CC arthritis.
 SO Sequence 1130 AA;

Db 797 gpgtlrarg 805 48.5%; SCORE 48; DB 14; Length 1130;
 Query 1 OSQTMRTTH 9 55.6%; Pred. No. 2.29e+02;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

RESULT 13
 ID R81569 standard; Protein; 1130 AA.
 AC R81569;
 DT 22-MAY-1996 (first entry)
 KW Class II transactivator.
 DE CIITA: class II transactivator; transcription; MHC;
 KM major histocompatibility complex class II; autoimmune disease;
 therapy.
 OS Homo sapiens.
 FH Key
 FT domain 26..352
 FT /note="transcription activation domain"
 FT 301..1130
 FT /note="interaction domain"
 FI domain
 FI MO9606107-A1.
 PN 29-FEB-1996.
 PD 22-AUG-1995; U10691.
 PF 24-AUG-1994; US-295502.
 PR (HARD) HARVARD COLLEGE.
 PA Douhan J, Glimcher LH, Zhou H;
 PI WPI; 96-151325/15.
 DR N-PSDB; R18028.
 PT Methods which inhibit transcription activation by CIITA - causes
 PT inhibition of MHC class II gene expression, used in therapy of
 PT auto-immune disorders
 PS Example 1: Page 36-41; 64pp; English.
 CC A genomic DNA sequence (R18028) codes for the class II
 CC transactivator, CIITA (R81569), a protein essential for activation
 CC of transcription of MHC class II genes. The CIITA transcription
 CC activation domain provides useful information for identifying cpds.
 CC which inhibit CIITA-dependent transcription. Such cpds. are

```

CC potential autoimmune disease therapeutics by virtue of their ability
CC to inhibit transcription of the MHC class II genes. A second portion
CC of CIITA, the interaction domain, mediates interaction of CIITA
CC and its target protein and activates MHC class II promoters.
CC Ccids. that inhibit this binding are also potential autoimmune
CC disease therapeutics.
SQ Sequence 1130 AA;

Db      797 gpgtlrarg 805
Qy      1 QSGTMRTRH 9

Query Match          48.5%; Score 48; DB 16; Length 1130;
Best Local Similarity 55.6%; Pred. No. 2.29e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

RESULT 14
ID W80311 standard; Protein; 1207 AA.
AC W80311;
DT 01-FEB-1999 (first entry)
DE CIITA (class II transactivator) type I protein.
KW CIITA gene; class II transactivator; MHC class II molecule;
KW Interferon-gamma; interleukin-4; vaccine; cancer treatment.
OS Homo sapiens.
PN EP-874049-A1.
PD 28-Oct-1998.
PF 21-Apr-1998; 400968.
PR 22-Apr-1997; FR-004954.
PA (TRGE ) TRANSGENE SA.
PI Mach B;
DR WP1; 98-559115/48.
PR DNA encoding MHC class II trans-activator polypeptide(s) - useful
PR e.g. as primers for enzymatic amplification, as detection probes or
PR as inhibitors of expression of genes encoding MHC in dendritic cells
PS Claim 10: Pages 53-57; 86pp; French.
CC The present sequence represents a type I CIITA (class II transactivator)
CC protein. The products can be used to treat diseases for which
CC enhanced expression of genes coding for MHC class II molecules is
CC desired, especially where the enhanced expression is desired in
CC dendritic cells or after induction by a cytokine, especially
CC interferon-gamma or interleukin-4. Inhibitors of the gene or protein
CC can be used to treat diseases for which reduced expression of genes
CC coding for MHC class II molecules is desired or can be used as vaccines
CC especially for cancer treatment.
SQ Sequence 1207 AA;

Query Match          48.5%; Score 48; DB 37; Length 1207;
Best Local Similarity 55.6%; Pred. No. 2.29e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db      874 gpgtlrarg 882
Qy      1 QSGTMRTRH 9

Query Match          48.5%; Score 48; DB 37; Length 1207;
Best Local Similarity 55.6%; Pred. No. 2.29e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

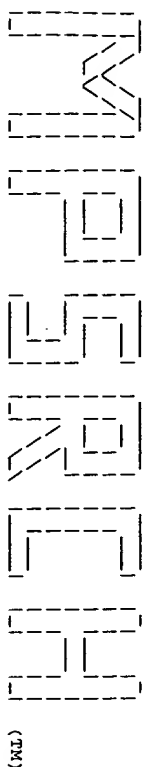
RESULT 15
ID W73510 standard; Protein; 483 AA.
AC W73510;
DT 02-MAR-1999 (first entry)
DE Alpha-amylase protein variant M197T.
KW PCR primer; alpha- amylase; variant; granular starch slurry;
KW oxidative stability.
OS Bacillus licheniformis.
PN US5649549-A-.
PD 15-DEC-1998.
PF 06-JUN-1995; 468698.
PR 10-FEB-1994; US-194664.
PR 11-FEB-1993; US-016395.
PR 06-JUN-1995; US-468698.
PA (GEMV ) GENENCOR INT.
PI Barnett CC, Mitchinson C, Power SD, Requadt CA,
PI Solheim LP;
```

DR WPI; 99-069736/06.
PT Starch liquefaction - using mutant Bacillus alpha-amylase
PS Disclosure: Fig 4a: 56pp; English.
CC This sequence represents the M197T variant of the Bacillus
CC licheniformis alpha-amylase. Variants of the protein can be used in
CC the method of the invention, for liquefying a granular starch slurry from
CC a wet or dry milling process at a pH of 4 to less than 6, that comprises
CC adding a mutant Bacillus alpha-amylase to the slurry, optionally adding
CC an antioxidant, and reacting the slurry until the starch is liquefied.
CC The alpha-amylase has a substitution of Thr, Leu, Asn or Asp for a Met
CC residue corresponding to M15 in B. licheniformis alpha-amylase.
CC Alternatively the alpha-amylase has a substitution of Leu or Ala for a
CC residue corresponding to M197 in B. licheniformis alpha amylase. Mutants
CC such as M197L and M15L have enhanced oxidative stability at pH 5.
SQ Sequence 483 AA;

Query Match 47.58; Score 47; DB 38; Length 483;
Best Local Similarity 53.8%; Pred. No. 2.87e+02;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 69 qkgtvrtkygkqg 81
| | | | | | | | | |
QY 1 QSGTMRTRHSTGG 13

Search completed: Fri Jun 11 17:54:36 1999
Job time : 106 secs.



Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mparch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Jun 11 17:57:22 1999; Maspar time 4.34 Seconds
Tabular output not generated. 138.367 Million cell updates/sec

Title: >US-08-991-628-7
Description: (1-15) from US08991628.pep
Perfect Score: 99
Sequence: 1 QSGTMRFRHSTGTN 15

Scoring table: PAM 150
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r60
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 23.847; Variance 29.890; scale 0.798

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	99	100.0	999	1	IGHUG3	desmoglein 3 precursor	1.83e-11
2	57	57.6	453	2	S14666	keratin 10, type I, c	3.92e-01
3	57	57.6	1597	1	BVFEFL	sol protein, large sp	3.92e-01
4	56	56.6	478	2	S33886	DNA-directed RNA poly	6.39e-01
5	55	55.6	534	2	A44991	protein-tyrosine kina	1.04e-00
6	55	55.6	534	2	S33568	protein-tyrosine kina	1.04e-00
7	55	55.6	537	2	A43806	protein-tyrosine kina	1.04e-00
8	55	55.6	537	2	S15192	p59(Xfyn) - Xiphophor	1.04e-00
9	55	55.6	537	1	TVHUSY	protein-tyrosine kina	1.04e-00
10	53	53.5	160	2	T03365	gene m3 protein - lac	2.59e+00
11	51	51.5	475	2	JN0578	p-aminobenzoic acid s	6.80e+00
12	51	51.5	481	2	S04605	glycinin G3 - soybean	6.80e+00
13	51	51.5	484	2	S11003	glycinin G3 precursor	6.80e+00
14	51	51.5	495	1	FMSYSG	glycinin chain Alabx	6.80e+00
15	51	51.5	495	2	S10851	glycinin G1 precursor	6.80e+00
16	50	50.5	286	2	JE0417	aminoglycoside-N-acet	1.07e+01
17	50	50.5	572	1	HNNZB3	hemagglutinin-neuram	1.07e+01
18	50	50.5	2318	1	S45306	notch 3 protein - mou	1.07e+01
19	49	49.5	140	2	A25979	hypothetical protein	1.68e+01
20	49	49.5	368	2	AG0854	MHC class I histocomp	1.68e+01
21	49	49.5	368	2	I49712	H-2K-s - mouse	1.68e+01
22	49	49.5	368	2	I49713	H-2K-sml - mouse	1.68e+01
23	48	48.5	63	2	G65026	hypothetical protein	2.61e+01

24	48	48.5	172	2	JE0130	scyalone dehydratase	2.61e+01
25	48	48.5	291	2	S29292	transcription factor	2.61e+01
26	48	48.5	462	2	A40552	bindin fertilization	2.61e+01
27	48	48.5	471	2	S08333	exo-alpha-galactase	2.61e+01
28	48	48.5	472	2	A70951	hypothetical protein	2.61e+01
29	48	48.5	510	1	WZBE47	protein-serine/threon	2.61e+01
30	48	48.5	729	2	S36605	nitrate reductase (EC	2.61e+01
31	48	48.5	858	2	S54119	phosphoprotein - flss	2.61e+01
32	48	48.5	907	2	A57087	cell division control	2.61e+01
33	48	48.5	1053	1	QYXC	phosphoenolpyruvate c	2.61e+01
34	48	48.5	1130	1	A48843	MHC class II transact	2.61e+01
35	48	48.5	1201	1	ERBYA	coarmer complex alph	2.61e+01
36	48	48.5	1591	2	A54146	invasion-inducing pro	2.61e+01
37	48	48.5	1879	2	S74915	extracellular nucleas	2.61e+01
38	47	47.5	41	2	T03647	probable ribosomal pr	4.03e+01
39	47	47.5	416	2	A32947	flaggrin precursor -	4.03e+01
40	47	47.5	442	2	A36469	dam protein - fruit f	4.03e+01
41	47	47.5	489	2	S42048	nucleocapsid protein	4.03e+01
42	47	47.5	512	1	ALBSL	alpha-amylase (EC 3.2	4.03e+01
43	47	47.5	587	2	T00316	coxR-regulated lipopr	4.03e+01
44	47	47.5	1487	2	T02850	probable membrane pro	4.03e+01
45	47	47.5	1792	2	A57075	tensin - chicken (fra	4.03e+01

ALIGNMENTS

RESULT	ENTRY	TYPE	COMPLETE
1	IGHUG3	#type complete	
	desmoglein 3 precursor	- human	
	ALTERNATE_NAMES	pemphigus vulgaris antigen	
	ORGANISM	#formal_name Homo sapiens #common_name man	
	DATE	30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 18-Sep-1998	

ACCESSIONS	REFERENCE
A41088	Amagai, M.; Klaus-Kovtun, V.; Stanley, J.R.
A41088	Cell (1991) 67:869-877
	Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a disease of cell adhesion.
	#cross-references NID:92069753
	#accession A41088

GENETICS	
#gene	GDB:DSG3
#map_position	18q12.1-18q12.2
CLASSIFICATION	#superfamily cadherin; cadherin repeat homology
KEYWORDS	calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein

FEATURE	
1-23	#domain signal sequence #status predicted #label SIG\
24-49	#domain propeptide #status predicted #label PRO\
50-999	#product desmoglein homolog #status predicted #label MAF\

50-615	52-517	160-267	270-383	390-495	496-598	616-639	640-999	910-938	937-966	110,180,545
#domain extracellular #status predicted #label EXT\	#domain extracellular repeat homology #label CR1\	#domain cadherin repeat homology #label CR2\	#domain cadherin repeat homology #label CR3\	#domain cadherin repeat homology #label CR4\	#domain cadherin repeat homology #label CR5\	#domain transmembrane #status predicted #label TM\	#domain intracellular #status predicted #label INT\	#domain desmoglein repeat #label DSG2\	#binding_site cardohydrate (Asn) (covalent) #status predicted	#length 999 #molecular-weight 107502 #checksum 8311

SUMMARY	
Query Match	100.0%; Score 99; DB 1; Length 999;
Best Local Similarity	100.0%; Pred. No. 1.83e-11;
Matches	15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 762 QSGMTRHSTGCTN 776
 |||||
 QY 1 QSGMTRHSTGCTN 15

RESULT 2
 ENTRY S14666 #type fragment
 TITLE keratin 10, type I, cytoskeletal (clone IIB5) - human (fragment)

ALTERNATE_NAMES cytokeratin 10
 ORGANISM #formal_name Homo sapiens #common_name man
 DATE 30-Sep-1991 #sequence_revision 31-Dec-1991 #text_change 05-Dec-1998

ACCESSIONS S14666
 REFERENCE S14666
 #authors Darmon, M.Y.; Semat, A.; Darmon, M.C.; Vasseur, M.
 #journal Mol. Biol. Rep. (1987) 12:277-283
 #title Sequence of a cDNA encoding human keratin No 10 selected according to structural homologues of keratins and their tissue-specific expression.

#cross-references M01D:88122104
 #accession S14666
 #molecule_type mRNA
 #residues 1-453 #label DARI
 #cross-references EMBL:M19156
 #note the sequence from Fig. 3 is inconsistent with the nucleotide sequence from Fig. 2 in having additional residues Ile-Lys-Ile-Arg-Leu after 313-Leu

REFERENCE S14667
 #authors Darmon, M.Y.; Semat, A.; Darmon, M.C.; Vasseur, M.
 #submission submitted to the EMBL Data Library, May 1988
 #accession S14669
 #molecule_type mRNA
 #residues 1-439, 'S', 441-453 #label DAR2
 #cross-references EMBL:M19156
 #superfamily cytoskeletal keratin
 #keywords coiled coil; heptad repeat; intermediate filament

FEATURE
 1-16 #domain head (fragment) #status predicted #label HRA\
 17-327 #domain helical rod #status predicted #label ROD\
 328-453 #domain tail #status predicted #label TAI
 #length 453 #checksum 1314

SUMMARY
 Query Match 57.6%; Score 57; DB 2; Length 453;
 Best Local Similarity 50.0%; Pred. No. 3,92e-01;
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 408 GGGLRHSGGGSS 421
 |||||
 QY 2 SGMTRHSTGCTN 15

RESULT 3
 ENTRY BVFFSL #type complete
 TITLE sol protein, large splice form - fruit fly (Drosophila melanogaster)
 ORGANISM #formal_name Drosophila melanogaster
 DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 18-Sep-1998

ACCESSIONS A41146
 REFERENCE A41146
 #authors Delaney, S.J.; Hayward, D.C.; Barleben, F.; Fischbach, K.F.; Mikoy, G.L.G.
 #journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:7214-7218
 #title Molecular cloning and analysis of small optic lobes, a structural brain gene of Drosophila melanogaster.

#cross-references M01D:91334436
 #accession A41146
 #molecule_type mRNA
 #residues 1-1597 #label DEL
 #cross-references GB:M64084; M1D:92760822; P1D:9158483

COMMENT The sol (small optic lobes) mutation eliminates certain classes of

columnar neurons.
 COMMENT An alternate splice form of 395 amino acids is observed, in which the first 393 are identical to the large sol protein.

GENETICS
 #gene sol
 #cross-references FlyBase:FBgn0003464
 #map_position 19F4
 CLASSIFICATION #superfamily sol protein; calpain catalytic domain homology
 KEYWORDS #feature alternative splicing; brain; zinc finger

FEATURE
 12-29 #region zinc finger CCCC motif\
 141-158 #region zinc finger CCCC motif\
 649-667 #region zinc finger CCCC motif\
 673-702 #region glutamine/histidine-rich\
 713-730 #region zinc finger CCCC motif\
 754-771 #region zinc finger CCCC motif\
 936-953 #region zinc finger CCCC motif\
 1047-1307 #domain calpain catalytic domain homology #label CALP\
 1082,1248,1268 #active_site Cys, His, Asn #status predicted
 #length 1597 #molecular_weight 174713 #checksum 8253

SUMMARY
 Query Match 57.6%; Score 57; DB 1; Length 1597;
 Best Local Similarity 61.5%; Pred. No. 3,92e-01;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 873 SGAPRHRSTGCS 885
 |||||
 QY 2 SGMTRHSTGCT 14

RESULT 4
 ENTRY S33886 #type fragment
 TITLE DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain
 ORGANISM Rp1 - Euploes octocarinatus (SGC9) (fragment)
 #formal_name Euploes octocarinatus
 DATE 08-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 12-Sep-1997

ACCESSIONS S33886; S70414; S73103
 REFERENCE S33884
 #authors Kaufmann, J.; Klein, A.
 #journal Nucleic Acids Res. (1992) 20:4445-4450
 #title Gene dosage as a possible major determinant for equal expression levels of genes encoding RNA polymerase subunits in the hypotrichous ciliate Euploes octocarinatus.

#cross-references M01D:93027138
 #accession S33886
 #status translation not shown
 #molecule_type DNA
 #residues 1-478 #label KAU

REFERENCE S70412
 #authors Kaufmann, J.; Florian, V.; Klein, A.
 #journal Nucleic Acids Res. (1992) 20:5985-5989
 #title TGA cysteine codons and intron sequences in conserved and nonconserved positions are found in macronuclear RNA polymerase genes of Euploes octocarinatus.

#cross-references M01D:93096587
 #accession S70414
 #molecule_type DNA
 #residues 1-124 #label KAU
 #cross-references EMBL:X66452

GENETICS
 #gene RP1
 #genetic_code SGC9
 #introns 30/3; 74/1
 CLASSIFICATION #superfamily human DNA-directed RNA polymerase II largest chain
 KEYWORDS DNA binding; nucleotidyltransferase; tandem repeat; transcription; zinc finger
 #length 478 #checksum 3090

SUMMARY
 Query Match 56.6%; Score 56; DB 2; Length 478;
 Best Local Similarity 53.8%; Pred. No. 6,39e-01;

Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 301 QAGQVSRHKSXG 313
 1 QSGTMRTRHSTG 13

RESULT 5
 ENTRY A44991 #type complete
 TITLE protein-tyrosine kinase (EC 2.7.1.112) fyn - mouse
 ALTERNATE_NAMES kinase-related transforming protein (fyn)
 ORGANISM #formal_name Mus musculus #common_name house mouse
 DATE 03-Jun-1993 #sequence_revision 30-Sep-1993 #text_change 08-Sep-1997

ACCESSIONS A44991
 REFERENCE A44991
 #authors Cooke, M.P.; Perlmuter, R.M.
 #journal New Biol. (1989) 1:66-74
 #title Expression of a novel form of the fyn proto-oncogene in hematopoietic cells.
 #cross-references MUID:91175680
 #accession A44991
 #status Preliminary
 #molecule_type mRNA
 #residues 1-534 #label COO
 #cross-references GB:M27266; NID:9193357; PID:9309241
 #note In the authors' translation an additional residue Leu was shown after Lys, for residue 459

CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
 #homology SH2 homology; SH3 homology
 #organism ATP; autophosphorylation; phosphoprotein; phosphotransferase; tyrosine-specific protein kinase

FEATURE
 89-138 #domain SH3 homology #label SH3\
 149-246 #domain SH2 homology #label SH2\
 266-524 #domain protein kinase homology #label KIN\
 274-282 #region protein kinase ATP-binding motif
 #length 534 #molecular_weight 60057 #checksum 2574

SUMMARY
 Query Match 55.6%; Score 55; DB 2; Length 534;
 Best Local Similarity 57.1%; Pred. No. 1.04e+00;
 Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 72 HTGTLRTRGCTGVT 85
 1 QSGTMRTRHSTG 14

RESULT 6
 ENTRY S33568 #type complete
 TITLE protein-tyrosine kinase (EC 2.7.1.112) fyn - chicken
 ORGANISM #formal_name Gallus gallus #common_name chicken
 DATE 31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change 08-Sep-1997

ACCESSIONS S33568; S36351; S20807
 REFERENCE S33568
 #authors Sudol, M.; Greulich, H.; Newman, L.; Sarkar, A.; Sukegawa, J.; Yamamoto, T.
 #journal Oncogene (1993) 8:823-831
 #title A novel yes-related kinase, Yrk, is expressed at elevated levels in neural and hematopoietic tissues.
 #cross-references MUID:93205395
 #accession S33568
 #molecule_type mRNA
 #residues 1-534 #label SUD1
 #cross-references EMBL:X52841
 REFERENCE S36351
 #authors Sudol, M.
 #submission submitted to the EMBL Data Library, May 1990
 #accession S36351
 #molecule_type mRNA
 #residues 1-348, 1-350-534 #label SUD2
 #cross-references EMBL:X52841; NID:962861; PID:962862

GENETICS
 #gene fyn
 CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
 #homology SH2 homology; SH3 homology
 #organism ATP; phosphoprotein; phosphotransferase; proto-oncogene; transforming protein; tyrosine-specific protein kinase

FEATURE
 89-138 #domain SH3 homology #label SH3\
 149-246 #domain SH2 homology #label SH2\
 266-524 #domain protein kinase homology #label KIN\
 274-282 #region protein kinase ATP-binding motif
 296 #binding_site ATP (Lys) #status predicted
 417-528 #binding_site phosphate (Tyr) (covalent) #status predicted

SUMMARY
 Query Match 55.6%; Score 55; DB 2; Length 534;
 Best Local Similarity 57.1%; Pred. No. 1.04e+00;
 Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 72 HTGTLRTRGCTGVT 85
 1 QSGTMRTRHSTG 14

RESULT 7
 ENTRY A43806 #type complete
 TITLE protein-tyrosine kinase (EC 2.7.1.112) fyn - African clawed frog
 ORGANISM #formal_name Xenopus laevis #common_name African clawed frog
 DATE 30-Jan-1993 #sequence_revision 30-Jan-1993 #text_change 31-Oct-1997

ACCESSIONS A43806
 REFERENCE A43806
 #authors Steele, R.E.; Deng, J.C.; Ghosh, C.R.; Fero, J.B.
 #journal Oncogene (1990) 5:369-376
 #title Structure and expression of fyn genes in Xenopus laevis.
 #cross-references MUID:90191723
 #accession A43806
 #status Preliminary
 #molecule_type mRNA
 #residues 1-537 #label STE
 #cross-references GB:X52188; GB:X52189; NID:964701; PID:964702

CLASSIFICATION #superfamily protein-tyrosine kinase src; protein kinase
 #homology SH2 homology; SH3 homology
 #organism ATP; autophosphorylation; phosphoprotein; phosphotransferase; tyrosine-specific protein kinase

FEATURE
 89-138 #domain SH3 homology #label SH3\
 149-246 #domain SH2 homology #label SH2\
 269-527 #domain protein kinase homology #label KIN\
 277-285 #region protein kinase ATP-binding motif
 #length 537 #molecular_weight 60446 #checksum 2563

SUMMARY
 Query Match 55.6%; Score 55; DB 2; Length 537;
 Best Local Similarity 57.1%; Pred. No. 1.04e+00;
 Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 72 HTGTLRTRGCTGVT 85
 1 QSGTMRTRHSTG 14

RESULT 8
 ENTRY I51592 #type complete
 TITLE p59(Xfyn) - xiphophorus helleri
 ORGANISM #formal_name Xiphophorus helleri
 DATE 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 12-Feb-1999

ACCESSIONS I51592
 REFERENCE I51592
 #authors Hannig, G.; Ottillie, S.; Scharf, M.
 #journal Oncogene (1991) 6:361-369

OC TETRAODONTIFORMES; TETRAODONTIOIDEI; TETRAODONTIDAE; FUGU.

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96382517.
RA POUTER R.T.M., BOUTER M.I.;
RT "A retrotransposon family from the pufferfish (fugu) *Fugu rubripes*,"
RL GENE 215:241-249(1998).
DR EMBL: AF030381; G3510505; -
KM POLYPROTEIN.
FT NON TER.
SQ SEQUENCE 1187 AA; 132431 MW; 7AEETFO1 CRC32;

Query Match 56.9%; Score 62; DB 13; Length 1187;
Best Local Similarity 60.0%; Pred. No. 8,75e+01;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

DB 1055 SSRKLNPRDIPPTI 1069
QY 1 SARTLNRRYTGPTF 15

RESULT 3
ID 024780 PRELIMINARY; PRT; 980 AA.
AC 024780;
DT 01-JAN-1998 (TREMREL. 05, CREATED)
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE CELLODEXTRIN PHOSPHORYLASE.
GN CDP.
OS CLOSTRIDIUM THERMOCCELLUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC27405;
RA KAMAGUCHI T., IKUCHI Y., TSUTSUMI N., KAN A., SUMITANI J., ARAI M.,
RL J. FERMENT. BIOENG. 85:144-149(1998).
DR EMBL: AB006822; D1022940; -
SQ SEQUENCE 980 AA; 111183 MW; ABBAE30C CRC32;

Query Match 56.0%; Score 61; DB 2; Length 980;
Best Local Similarity 60.0%; Pred. No. 1,35e+00;
Matches 9; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

DB 713 ARVLINRYKDSYTY 727
QY 2 ARTLNRRYTGPTF 15

RESULT 4
ID 081508 PRELIMINARY; PRT; 973 AA.
AC 081508;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE 17M2.4.4 PROTEIN.
GN 17M2.4.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STEPHOPHYTA; EMBRIOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHITES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPRALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA HARMON G., LANGSTON Y., STONKING T., DRONE K., AMES M.;
RT "The sequence of *Arabidopsis thaliana* 17M2.4,"
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF030381; G3510505; -

Query Match 54.1%; Score 59; DB 2; Length 163;
Best Local Similarity 87.5%; Pred. No. 3,20e+00;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA WATERSTON R.;
RT SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF077408; G3319363; -
SQ SEQUENCE 973 AA; 112369 MW; F6086EE2 CRC32;

Query Match 55.0%; Score 60; DB 10; Length 973;
Best Local Similarity 53.8%; Pred. No. 2,09e+00;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

DB 826 SRKLSPRYVGPY 838
QY 1 SARTLNRRYTGPT 13

RESULT 5
ID 053240 PRELIMINARY; PRT; 163 AA.
AC 053240;
DT 01-JUN-1998 (TREMREL. 06, CREATED)
DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 18.2 KD PROTEIN.
GN MTVO12.05.
OS MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; CORINEBACTERIALES; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA OLIVER K., HARRIS D.;
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA BARRELL B.G., RAJANDREAM M.A., PARKHILL J., COLE S.T.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE: 96181548.
RA PHILIP W.J., POULET S., EIGMEIER K., PASCOPELLA L.,
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RT "An integrated map of the genome of the tubercle bacillus,
RT Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium
RT leprae.";
RT PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).
DR EMBL: AL021287; E1237736; -
KM HYPOTHETICAL PROTEIN.
SQ SEQUENCE 163 AA; 18204 MW; D33CE4BB CRC32;

Query Match 54.1%; Score 59; DB 2; Length 163;
Best Local Similarity 87.5%; Pred. No. 3,20e+00;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 111 ERTGPT 118
QY 7 NRYTGPT 14

RESULT 6
ID 059397 PRELIMINARY; PRT; 330 AA.
AC 059397;
DT 01-AUG-1998 (TREMREL. 07, CREATED)
DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)
DE 330AA LONG HYPOTHETICAL N-ACETYL-GAMMA-GLUTAMYL-PHOSPHATE REDUCTASE.
GN PH1720.
OS PYROCOCCUS HORIKOSHII.
OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCCUS.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-OT3;
RX MEDLINE; 98344137.
RA KANABAYASTI Y., SAMADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOVAMA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTURA R., NAKAZAWA H., TAKAKIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT "Complete Sequence and Gene Organization of the genome of a
RT Hyper-thermophilic Archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA RES. 5:55-76(1998).
DR EMBL: AF000007; D103177;
SQ SEQUENCE 330 AA; 37168 MW; D80618ED CRC32;

Query Match 54.1%; Score 59; DB 1; Length 330;
Best Local Similarity 53.8%; Pred. No. 3.20e+00;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 50 RGLNRTSDYNF 62
| | | | |
OY 3 RLLNRRYTGPTF 15

RESULT 7
ID 050378 PRELIMINARY; PRT; 360 AA.
AC 050378;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE FERRIC EXOCHELIN BIOSYNTHESIS.
GN FXBA.
OS MYCOBACTERIUM SMEGMATIS.
OC BACTERIA: FIRMICUTES: ACTINOBACTERIA: ACTINOBACTERIIDE;
OC ACTINOMYCETALES: CORNYNBACTERIINAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MC6;
RX MEDLINE; 95191405.
RA FISS E.H.;
RT "Identification of genes involved in the sequestration of iron in
RT mycobacteria: the ferric exochelatin biosynthetic and uptake
RT pathways.";
RL MOL. MICROBIOL. 14:557-569(1994).
DR EMBL: U10425; G595404;
DR PRAM: PF00551; formyl.transf. 1.
SQ SEQUENCE 360 AA; 41130 MW; A81D7F3B CRC32;

Query Match 54.1%; Score 59; DB 2; Length 360;
Best Local Similarity 38.5%; Pred. No. 3.20e+00;
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 241 RAISDPYNAVY 253
| | | | |
OY 3 RLLNRRYTGPTF 15

RESULT 8
ID 066695 PRELIMINARY; PRT; 531 AA.
AC 066695;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHEICAL 62.8 KD PROTEIN.
GN AC 372.
OS AQUIFEX AEOLICUS.
OC BACTERIA: AQUIFICALES; AQUIFICACEAE; AQUIFEX.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RX MEDLINE; 98196666.
DR DICKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., ADJAY M., HUBER R.,

RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL NATURE 392:353-358(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEER R., SNEAD M.A., KELLER M., ADJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AE000686; G2983043;
KM HYPOTHEICAL PROTEIN.
SQ SEQUENCE 531 AA; 62830 MW; 585EBE7E CRC32;

Query Match 54.1%; Score 59; DB 2; Length 531;
Best Local Similarity 66.7%; Pred. No. 3.20e+00;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 80 TLKRLNRYPTF 91
| | | | |
OY 4 TLNRRYTGPTF 15

RESULT 9
ID 001124 PRELIMINARY; PRT; 211 AA.
AC 001124;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COAT PROTEIN VPI (FRAGMENT).
GN ID.
OS FOOT-AND-MOUTH DISEASE VIRUS (STRAIN C5 ARGENTINA/65)
OC (APHTHOVIRUS C).
OC VIRUSES: SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC APHTHOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92260626.
RA MARTINEZ M.A., DOPAZO J., HERNANDEZ J., MATEU M.G., SOBRINO F.,
RA DOMINGO E., KNOWLES N.J.;
RT "Evolution of the capsid protein genes of foot-and-mouth disease
RT virus: antigenic variation without accumulation of amino acid
RT substitutions over six decades.";
RL J. VIROL. 66:3557-3565(1992).
CC -1- THE COAT PROTEIN VPI CONTAINS THE MAIN ANTIGENIC DETERMINANTS OF
CC THE VIRION. THEREFORE, CHANGES IN ITS SEQUENCE MUST BE
CC RESPONSIBLE FOR THE HIGH ANTIGENIC VARIABILITY OF THE VIRUS.
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
DR EMBL: M90377; G210394; -
DR PRAM: PF00073; fhv. 1.
KM COAT PROTEIN. 1
FT NON_TER 1
FT DOMAIN 134 157
FT DOMAIN 188 211
FT NON_TER 211 211
SQ SEQUENCE 211 AA; 23028 MW; B1B47466 CRC32;

Query Match 53.2%; Score 58; DB 14; Length 211;
Best Local Similarity 53.8%; Pred. No. 4.88e+00;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Db 125 RVLATRYGTGTTY 137
| | | | |
OY 3 RLLNRRYTGPTF 15

RESULT 10
ID 029749 PRELIMINARY; PRT; 332 AA.
AC 029749;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)

```

DE 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
GN 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
OC NITRATE REDUCTASE, GAMMA SUBUNIT, PUTATIVE.
OS AF0501.
OS ARCHAEOGLOBUS FULGIDUS.
OS ARCHAEEA; EURYARCHAEOTA; ARCHAEOGLOBALES; ARCHAEOGLOBACEAE;
OC ARCHAEOGLOBUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VC-16 / DSM 4304 / ATCC 49538;
RX MEDLINE; 98049343.
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTTACH P., KATNE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic,
RT sulphate-reducing archaeon Archaeoglobus fulgidus.";
RL NATURE 390:364-370(1997).
DR EMBL; AE001069; G2650124; -.
DR TIGR; AF0501; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 332 AA; 38386 MW; 6844D7B8 CRC32;

Query Match 53.2%; Score 58; DB 1; Length 332;
Best Local Similarity 50.0%; Pred. No. 4.88e+00;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 53 RTIYDFRSPYT 64
QY 3 RTLNRYTGPYT 14
||| :||: |||

RESULT 11
AC Q44145; PRELIMINARY; PRT; 444 AA.
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE NIFN.
GN NIFN.
US ANABAENA 7120.
OC BACTERIA; CYANOBACTERIA; NOSTOCALES; NOSTOCACEAE; ANABAENA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC 7120;
J., SCAPPINO L.A., HASELKORN R.;
(JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
15; G1185622; -.
148; oxidized nitro: 1.
444 AA; 48332 MW; 9C962756 CRC32;

Query Match 53.2%; Score 58; DB 2; Length 444;
Best Local Similarity 46.2%; Pred. No. 4.88e+00;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 245 AARLEDFRNIPY 257
QY 1 SARTLNRYTGPY 13
||| :||: |||

RESULT 12
ID Q17381; PRELIMINARY; PRT; 506 AA.
AC Q17381;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
GN 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
OC NITRATE REDUCTASE, GAMMA SUBUNIT, PUTATIVE.
OS AF0501.
OS ARCHAEOGLOBUS FULGIDUS.
OS ARCHAEEA; EURYARCHAEOTA; ARCHAEOGLOBALES; ARCHAEOGLOBACEAE;
OC ARCHAEOGLOBUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VC-16 / DSM 4304 / ATCC 49538;
RX MEDLINE; 98049343.
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTTACH P., KATNE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic,
RT sulphate-reducing archaeon Archaeoglobus fulgidus.";
RL NATURE 390:364-370(1997).
DR EMBL; AE001069; G2650124; -.
DR TIGR; AF0501; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 332 AA; 38386 MW; 6844D7B8 CRC32;

Query Match 53.2%; Score 58; DB 5; Length 506;
Best Local Similarity 50.0%; Pred. No. 4.88e+00;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 443 TLNGQYASPLY 454
QY 4 TLNNRYTGPYT 15
||| :||: |||

RESULT 13
ID Q07117; PRELIMINARY; PRT; 770 AA.
AC Q07117;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 87.2 KD PROTEIN.
OS BARLEY STRIPE MOSAIC VIRUS (BSMV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; HORDEIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90066400.
RA KOZLOV YU.V., AFANAS'EV B.N., RUPASOV V.V.; GOLOVA YU.B.,
RA KULAYEVA O.I., DOLYA V.V., ATABEKOV I.G.; BAEV A.A.;
RT "Primary structure of RNA 3 of barley stripe mosaic virus and its
RT variability.";
RL MOL. BIOL. MOSK. 23:1080-1090(1989).
DR EMBL; X52774; G297555; -.
DR PFAM; PF00978; RNA_dep_RNapol2; 1.
KW HYPOTHETICAL PROTEIN.
RP SEQUENCE 770 AA; 87176 MW; 34477E66 CRC32;

Query Match 53.2%; Score 58; DB 14; Length 770;
Best Local Similarity 50.0%; Pred. No. 4.88e+00;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 720 KALVNYRSPYS 731
QY 3 RTLNRYTGPYT 14
||| :||: |||

RESULT 14
ID Q76025; PRELIMINARY; PRT; 80 AA.
AC Q76025;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OC HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1).
OS VIRUSES; RETROVIRIDAE; LENTIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
```

```

#title      Conservation of structure and expression of the c-yes and fyn
#            genes in lower vertebrates.
#cross-references MUID:91187435
#accession   I51592
#status      preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues    1-537 #label HAN
#cross-references EMBL:X54971; NID:964481; PID:964482
GENETICS
#gene        Xfyn
#description #superfamily protein-tyrosine kinase src; protein kinase
#            homology; SH2 homology; SH3 homology
FEATURE
89-138      #domain SH3 homology #label SH3\
149-246     #domain SH2 homology #label SH2\
269-527     #domain protein kinase homology #label KIN
SUMMARY
#length 537 #molecular-weight 60447 #checksum 621
Query Match      55.68; Score 55; DB 2; Length 537;
Best Local Similarity 57.1%; Pred. No. 1.04e+00;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
Db 72 HTGTLRTRGGTGT 85
QY 1 QSGTMRHSTGTT 14
      :::::|::|::|
RESULT 9
ENTRY   TVHUSY      #type complete
TITLE   protein-tyrosine kinase (EC 2.7.1.112) fyn, splice form B -
        human
ALTERNATE_NAMES  fynB; p59(fyn); protein-tyrosine kinase slk; protein-tyrosine
ORGANISM  kinase syn; proto-oncogene fyn
#formal_name Homo sapiens #common_name man
DATE      31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change
22-May-1998
ACCESSIONS A24314; A25389; I59120; I79512
REFERENCE  A24314
#authors   Samba, K.; Nishizawa, M.; Miyajima, N.; Yoshida, M.C.;
           Sukegawa, J.; Yamanashi, Y.; Sasaki, M.; Yamamoto, T.;
           Toyoshima, K.
#journal   Proc. Natl. Acad. Sci. U.S.A. (1986) 83:5459-5463
#title     yes-related protooncogene, syn, belongs to the
           protein-tyrosine kinase family.
#cross-references MUID:86287278
#accession A24314
#molecule_type mRNA
#residues 1-537 #label SEM
#cross-references GB:M14333; NID:9181171; PID:9181172
REFERENCE  A25389
#authors   Kawakami, T.; Pennington, C.Y.; Robbins, K.C.
#journal   Mol. Cell. Biol. (1986) 6:4195-4201
#title     Isolation and oncogenic potential of a novel human src-like
           gene.
#cross-references MUID:87089775
#accession A25389
#molecule_type mRNA
#residues 1-183; 'S', 185-436; 'R', 438-537 #label KAW1
#cross-references GB:M14676; NID:9338227; PID:9338228
#note     the authors translated the codon GAA for residue 265 as
           Gln and GGA for residue 278 as Gln
REFERENCE  I59120
#authors   Kawakami, T.; Kawakami, Y.; Aaronson, S.A.; Robbins, K.C.
#journal   Proc. Natl. Acad. Sci. U.S.A. (1988) 85:3870-3874
#title     Acquisition of transforming properties by fyn, a normal
           SRC-related human gene.
#cross-references MUID:88234523
#accession I59120
#status      translated from GB/EMBL/DBJ
#molecule_type DNA
#residues    514-522 #label KAW2
#cross-references GB:M20284; NID:9182842; PID:9182843
#experimental_source clone pFYN(C)-11; termination mutant p56(fyn)

```

```

#accession  I79512
#status      translated from GB/EMBL/DBJ
#molecule_type DNA
#residues    514-524; 'QRS', #label KAW3
#cross-references GB:M20285; NID:9182844; PID:9182845
#experimental_source clone pFYN(C)-21; recombinant mutant p57(fyn)
REFERENCE  A58826
#authors     Peters, D.J.; McGrew, B.R.; Perron, D.C.; Liptak, L.M.;
           Laudano, A.P.
#journal     Oncogene (1990) 5:1313-1319
#title       In vivo phosphorylation and membrane association of the fyn
           proto-oncogene product in IM-9 human lymphoblasts.
#cross-references MUID:91016431
#contents    annotation; myristylation; Tyr-531 phosphorylation
REFERENCE  A51398
#authors     Noble, M.; Musacchio, A.; Saraste, M.; Wierenga, R.
#submission  submitted to the Brookhaven Protein Data Bank, May 1993
#cross-references PDB:1SHF
#contents    annotation; X-ray crystallography, 1.9 angstroms, residues
           84-142
REFERENCE  A58827
#authors     Noble, M.E.M.; Musacchio, A.; Saraste, M.; Courtneidge, S.A.;
           Wierenga, R.K.
#journal     EMBO J. (1993) 12:2617-2624
#title       Crystal structure of the SH3 domain in human Fyn; comparison
           of the three-dimensional structures of SH3 domains in
           tyrosine kinases and spectrin.
#cross-references MUID:93327750
#contents    annotation; X-ray crystallography, 1.9 angstroms
REFERENCE  A56633
#authors     Musacchio, A.; Saraste, M.; Wilmanns, M.
#submission  submitted to the Brookhaven Protein Data Bank, May 1995
#cross-references PDB:1FYN
#contents    annotation; X-ray crystallography, 2.3 angstroms, residues
           81-142
REFERENCE  A58828
#authors     Musacchio, A.; Saraste, M.; Wilmanns, M.
#journal     Nature Struct. Biol. (1994) 1:546-551
#title       High-resolution crystal structures of tyrosine kinase SH3
           domains complexed with proline-rich peptides.
#cross-references MUID:95393198
#contents    annotation; X-ray crystallography, 2.3 angstroms
REFERENCE  A66288
#authors     Morton, C.J.; Pugh, D.J.R.; Campbell, I.D.
#submission  submitted to the Brookhaven Protein Data Bank, April 1996
#cross-references PDB:1NYF
#contents    annotation; conformation by (1)H-NMR, residues 84-141
REFERENCE  A66289
#authors     Morton, C.J.; Pugh, D.J.R.; Campbell, I.D.
#submission  submitted to the Brookhaven Protein Data Bank, April 1996
#cross-references PDB:1NYG
#contents    annotation; conformation by (1)H-NMR, residues 84-141
GENETICS
#gene        GDB:FYN
#cross-references GDB:118797; OMIM:137025
#map_position 6q21-6q21
COMPLEX
FUNCTION
#description catalyzes the phosphorylation of a peptidyl tyrosine residue
           by ATP
CLASSIFICATION
#superfamily protein-tyrosine kinase src; protein kinase
#homology; SH2 homology; SH3 homology
#alternative splicing; ATP; autophosphorylation; blocked amino
#end; lipoprotein; monomer; myristylation; phosphoprotein;
#phosphotransferase; proto-oncogene; thiolester bond;
#transforming protein; tyrosine-specific protein kinase
KEYWORDS
#domain SH3 homology #label SH3\
#domain SH2 homology #label SH2\
#domain protein kinase homology #label KIN\
#region protein kinase ATP-binding motif\
#modified_site myristylated amino end (Gly) (in mature
#form) #status predicted\
FEATURE
89-138
149-246
269-527
277-285
2

```

```

9      3.6          #binding_site palmitate (Cys) (covalent) #status
                predicted\
299         299     #active_site lys #status predicted\
420         420     #binding_site phosphate (Tyr) (covalent) (by
                    #autophosphorylation) #status predicted\
531         531     #binding_site phosphate (Tyr) (covalent) #status
                    experimental
SUMMARY        #length 537 #molecular-weight 60762 #checksum 9747

Query Match    55.6%; Score 55; DB 1; Length 537;
Best Local Similarity 57.1%; Pred. No. 1.04e+00;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db            72 HTGTLRTRGCTGT 85
QY           1 QSGTRMTRHSTGT 14
              ::::::::::::::|
RESULT       10
ENTRY        T03369             #type complete
TITLE        gene m3 protein - Lactococcus phage b11170
ORGANISM     #formal_name Lactococcus phage b11170
DATE         24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
                24-Mar-1999
ACCESSIONS   T03369
REFERENCE     C14903
AUTHORS      Crute-Le Cocq, A.M.; Cesaselin, B.; Commissaire, J.; Abba, J.;
                Kyriakidis, S.; Chopin, M.C.
SUBMISSION   submitted to the EMBL Data Library, June 1997
DESCRIPTION   Sequence and organization of the lactococcal isometric b11170
                phage genome.
ACCESION     T03369
STATUS       preliminary; translated from GB/EMBL/DBJ
MOLECULE_TYPE DNA
RESIDUES     1-160 ##label CRU
CROSS-REFERENCES EMBL:AF009650; NID:g3282260; PID:g3282321
GENETICS     m3
SUMMARY      #length 160 #molecular-weight 18072 #checksum 3777

Query Match    53.5%; Score 53; DB 2; Length 160;
Best Local Similarity 60.0%; Pred. No. 2.69e+00;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db            90 GVMRNHRNIG 99
QY           3 GTWRTNRHSTG 12

RESULT       11
ENTRY        JN0578             #type complete
TITLE        P-Aminobenzoic acid synthetase (EC 4.1.3.-) B - Streptomyces
                lividans
ORGANISM     #formal_name Streptomyces lividans
                14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change
                21-Aug-1998
ACCESSIONS   JN0578
REFERENCE     JN0577
AUTHORS      Arlin, F.F.; Vining, L.C.
JOURNAL      Gene (1993) 126:129-133
TITLE        Organization of the genes encoding P-aminobenzoic acid
                synthetase from Streptomyces lividans 1326.
CROSS-REFERENCES MIMD:93231526
ACCESSION     JN0578
MOLECULE_TYPE DNA
RESIDUES     1-475 ##label ARH
CROSS-REFERENCES GB:M64859; NID:g153393; PID:g153394
COMMENT      This enzyme participates in p-aminobenzoic acid biosynthesis.
GENETICS
SUMMARY      PabB
CLASSIFICATION #superfamily anthranilate synthase component I
KEYWORDS       carbon-carbon lyase; oxo-acid-lyase; p-aminobenzoate
                biosynthesis

```

```

SUMMARY      #length 475  #molecular-weight 52042  #checksum 6682

Query Match  51.5%;  Score 51;  DB 2;  Length 475;
Best Local Similarity 42.9%;  Pred. No. 6,80e+00;
Matches 6;  Conservative 4;  Mismatches 4;  Indels 0;  Gaps 0;

Db 281 AGTRGPRAGDD 294
QY 2 SGTMRTRHSTGCTN 15

RESULT 12
ENTRY      S04605      #type complete
TITLE      glycyltin G3 - soybean
ORGANISM   #formal_name Glycine max #common_name soybean
DATE       28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change
           08-Sep-1997

ACCESSIONS
REFERENCE   S04605
            Cho, T.J.; Nielsen, N.C.
            Nucleic Acids Res. (1989) 17,4388
            #title      The glycyltin Gy(3) gene from soybean.
            #cross-references MIMD:89296500
            #accession   S04605
            #molecule_type DNA
            #residues    1-481 #label CHO
            #cross-references EMBL:X15123; NID:G18638; PID:G18639
            #experimental_source variety Dare

GENETICS
#gene       Gy3
#introns    96/1; 177/3; 352/3
CLASSIFICATION #superfamily glycyltin
SUMMARY      #length 481 #molecular-weight 54241 #checksum 2269

Query Match  51.5%;  Score 51;  DB 2;  Length 481;
Best Local Similarity 58.3%;  Pred. No. 6,80e+00;
Matches 7;  Conservative 2;  Mismatches 3;  Indels 0;  Gaps 0;

Db 304 TMRHRNIGOTS 315
QY 4 TMRTRHSTGCTN 15

RESULT 13
ENTRY      S11003      #type complete
TITLE      glycyltin G3 precursor - soybean
ORGANISM   #formal_name Glycine max #common_name soybean
DATE       21-Nov-1993 #sequence_revision 19-Jan-1996 #text_change
           03-May-1996

ACCESSIONS
REFERENCE   S11003
            Nielsen, N.C.; Dickinson, C.D.; Cho, T.J.; Thanh, V.H.;
            Scallan, B.J.; Fischer, R.L.; Sims, T.L.; Drews, G.N.;
            Goldberg, R.B.
            Plant Cell (1989) 1:313-328
            #title      Characterization of the glycyltin gene family in soybean.
            #cross-references MIMD:92393391
            #accession   S11003
            #molecule_type DNA
            #residues    1-484 #label NIE
            #experimental_source variety Dare

GENETICS
#gene       Gy3
CLASSIFICATION #superfamily glycyltin
KEYWORDS     storage protein
FEATURE
1-19         #domain signal sequence #status predicted #label SIG\
20-484       #product glycyltin G3 #status predicted #label MAT
SUMMARY      #length 484 #molecular-weight 54359 #checksum 8616

Query Match  51.5%;  Score 51;  DB 2;  Length 484;
Best Local Similarity 58.3%;  Pred. No. 6,80e+00;
Matches 7;  Conservative 2;  Mismatches 3;  Indels 0;  Gaps 0;

```

```

Db      304  TMRHRNIGOTS 315
      ||| ||| ||| |||
Oy      4  TMRHRSTGNTN 15

RESULT  14
ENTRY   FW5YIG2      #type complete
TITLE   glycinin chain AlaBx precursor - soybean
ALTERNATE_NAMES  115 globulin; glycinin AlaBb
ORGANISM #formal name glycinine max #common name soybean
DATE     31-Mar-1988 #sequence-revision 31-Mar-1988 #text-change
22-May-1998

ACCESSIONS
REFERENCE A23497; S10502
#authors  Negoro, T.; Momma, T.; Fukazawa, C.
#journal  Nucleic Acids Res. (1985) 13:6719-6731
#title    A cDNA clone encoding a glycinin A-1a subunit precursor of
soybean.
#cross-references MIMD:86041867
#accession  A23497
#molecule_type mRNA
#residues  1-495 #label NEG
#cross-references GB:X02985; NID:g18614; PID:g18615
#experimental_source cv. Boninori1
#note     the authors translated the codon AAC for residue 449 as
Lys
#note     because of current nomenclature ambiguities, the authors
choose to designate the basic chain as Bx

REFERENCE S10502
#authors  Kitamura, Y.; Arakura, M.; Itoh, Y.; Fukazawa, C.
#journal  Nucleic Acids Res. (1990) 18:4245
#title    The complete nucleotide sequence of soybean glycinin A2B1a
gene spanning to another glycinin gene AlaBb.
#cross-references MIMD:90332420
#accession  S10502
#status    preliminary; translation not shown
#molecule_type DNA
#residues  481-495 #label KIT
#cross-references EMBL:X53404; NID:g18522; PID:g18523
#note     The source of this protein was cotyledon tissue taken 38 days after
flowering.
COMMENT   By homology with the A2B1a component, residues 307-310, and 491-495
are removed from the precursor during posttranslational
processing.
CLASSIFICATION #superfamily glycinin
KEYWORDS      seed; storage protein
FEATURE
1-19          #domain signal sequence #status predicted #label SIG\
20-306        #product glycinin chain Ala #status predicted #label
G1A\
311-490        #product glycinin chain Bx #status predicted #label GLB\
107-317        #disulfide bonds #status predicted
SUMMARY       #length 495 #molecular-weight 55506 #checksum 8518

Query Match 51.5%; Score 51; DB 1; Length 495;
Best Local Similarity 58.3%; Pred. No. 6.80e+00;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db      318  TMRHRNIGOTS 329
      ||| ||| ||| |||
Oy      4  TMRHRSTGNTN 15

RESULT  15
ENTRY   S10851      #type complete
TITLE   glycinin G1 precursor - soybean
ALTERNATE_NAMES  glycinin AlaBx
CONTAINS  glycinin chain Ala: glycinin chain Bx
ORGANISM #formal name Glycine max #common name soybean
DATE     21-Nov-1993 #sequence-revision 19-Jan-1996 #text-change
08-Sep-1997
ACCESSIONS S10851; S04603; JS0015

```

```

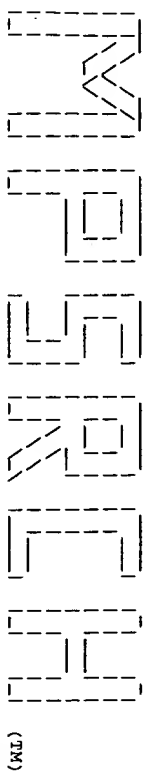
REFERENCE S10851
#authors  Nielsen, N.C.; Dickinson, C.D.; Cho, T.J.; Thanh, V.H.;
Scallion, B.J.; Fischer, R.L.; Sims, T.L.; Drews, G.N.;
Goldberg, R.B.
#journal  Plant Cell (1989) 1:313-328
#title    Characterization of the glycinin gene family in soybean.
#cross-references MIMD:92393391
#accession  S10851
#molecule_type DNA
#residues  1-495 #label NIE
#experimental_source variety Dare
REFERENCE S04603
#authors  Sims, T.L.; Goldberg, R.B.
#journal  Nucleic Acids Res. (1989) 17:4386
#title    The glycinin GY(1) gene from soybean.
#cross-references MIMD:89296498
#accession  S04603
#molecule_type DNA
#residues  1-495 #label SIM
#cross-references EMBL:X15121; NID:g18634; PID:g18635
#experimental_source variety Dare
REFERENCE JS0015
#authors  Utsunji, S.; Kohno, M.; Mori, T.; Kito, M.
#journal  J. Agric. Food Chem. (1987) 35:210-214
#title    An alternate cDNA encoding glycinin Ala Bx subunit.
#accession  JS0015
#molecule_type mRNA
#residues  1-495 #label UTS
#experimental_source var. Shirotsurunoko
GENETICS
#gene  Gyl
#introns  96/1; 180/3; 366/3
CLASSIFICATION #superfamily glycinin
KEYWORDS      storage protein
FEATURE
1-19          #domain signal sequence #status predicted #label SIG\
20-495        #product glycinin G1 #status predicted #label MAT1\
20-310        #product glycinin Ala chain #status predicted #label
MAT2\
311-495        #product glycinin Bx chain #status predicted #label MAT3
SUMMARY       #length 495 #molecular-weight 55706 #checksum 7917

Query Match 51.5%; Score 51; DB 2; Length 495;
Best Local Similarity 58.3%; Pred. No. 6.80e+00;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db      318  TMRHRNIGOTS 329
      ||| ||| ||| |||
Oy      4  TMRHRSTGNTN 15

Search completed: Fri Jun 11 17:58:45 1999
Job time : 83 secs.

```



Release 3.1A John F. Collins, Bioinformatics Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Jun 11 17:52:22 1999; Maspar time 2.45 Seconds
Tabular output not generated.

Title: >US-08-991-628-7
Description: (1-15) from US08991628.pep
Perfect Score: 99
Sequence: 1 OSGTMRTHSTGCTN 15

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-Processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 24.516; Variance 27.383; scale 0.895

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	99	100.0	999	1	DSG3_HUMAN DESMOGLEIN 3 PRECURSOR	4.72e-13
2	57	57.6	1597	1	SOL_DROME SMALL OPTIC LOBES PROT	1.22e-01
3	56	56.6	478	1	REPL_EUPOC DNA-DIRECTED RNA POLYM	2.09e-01
4	55	55.6	533	1	CHICK PROTO-ONCOGENE TYROSIN	3.54e-01
5	55	55.6	533	1	CHICK PROTO-ONCOGENE TYROSIN	3.54e-01
6	55	55.6	536	1	CHICK PROTO-ONCOGENE TYROSIN	3.54e-01
7	55	55.6	536	1	CHICK PROTO-ONCOGENE TYROSIN	3.54e-01
8	55	55.6	536	1	CHICK PROTO-ONCOGENE TYROSIN	3.54e-01
9	55	55.6	536	1	CHICK PROTO-ONCOGENE TYROSIN	3.54e-01
10	51	51.5	475	1	PABP_HUMAN T-COMPLEX PROTEIN 1, E	1.67e+00
11	51	51.5	481	1	GICL_SOYBN GLYCININ G1 PRECURSOR	2.76e+00
12	51	51.5	495	1	GICL_SOYBN GLYCININ G1 PRECURSOR	2.76e+00
13	50	50.5	120	1	IMMUNOGLOBULIN M1ACIDIAL	4.53e+00
14	50	50.5	286	1	HEXA_P13B HEMAGGLUTININ-NEURAMIN	4.53e+00
15	50	50.5	572	1	HEXA_P13B HEMAGGLUTININ-NEURAMIN	4.53e+00
16	50	50.5	2318	1	NEUROGENIC LOCUS NOTCH	4.53e+00
17	49	49.5	230	1	RL4_MYCIE 50S RIBOSOMAL PROTEIN	7.38e+00
18	48	48.5	172	1	SCID_MAGGR SCYTRALONE DEHYDRATASE	1.19e+01
19	48	48.5	291	1	T2EB_HUMAN TRANSCRIPTION INITIAT	1.19e+01
20	48	48.5	462	1	BIND_LYVA BINDIN PRECURSOR	1.19e+01
21	48	48.5	471	1	NEURAMINIDASE (EC 3.2.	1.19e+01
22	48	48.5	510	1	CATF_PSESY CATALASE PRECURSOR (EC	1.19e+01
23	48	48.5	510	1	KR2_VZVD PROBABLE SERINE/THREON	1.19e+01

24	48	48.5	729	1	NARB_SYNP7 NITRATE REDUCTASE (EC	1.19e+01
25	48	48.5	907	1	CC15_SCHPO CELL DIVISION CONTROL	1.19e+01
26	48	48.5	1053	1	CAPP_ANANI PHOSPHOENOLPYRUVATE CA	1.19e+01
27	48	48.5	1125	1	PHYA_POPTM PHYTOCHROME A	1.19e+01
28	48	48.5	1130	1	C2TA_HUMAN MHC CLASS II TRANSACTI	1.19e+01
29	48	48.5	1201	1	COPA_YEAST COTANOMER ALPHA SUBUNIT	1.19e+01
30	48	48.5	1591	1	T-LYMPHOMA INVASION AN	1.19e+01
31	48	48.5	1591	1	T-LYMPHOMA INVASION AN	1.19e+01
32	47	47.5	41	1	R58_MAIZE 40S RIBOSOMAL PROTEIN	1.91e+01
33	47	47.5	82	1	SE1B_ARATH PROTEIN TRANSPORT PROT	1.91e+01
34	47	47.5	370	1	MAF2_MOUSE TRANSCRIPTION FACTOR M	1.91e+01
35	47	47.5	389	1	ACD_MYCLE ACYL-COA DEHYDROGENASE	1.91e+01
36	47	47.5	416	1	FI1A_HUMAN FI1A-AMYLASE PRECURSOR (F	1.91e+01
37	47	47.5	469	1	NIFN_RHISN NITROGENASE IRON-MOLYB	1.91e+01
38	47	47.5	470	1	NRAM_IARUE NEURAMINIDASE (EC 3.2.	1.91e+01
39	47	47.5	470	1	NRAM_IARUE NEURAMINIDASE (EC 3.2.	1.91e+01
40	47	47.5	470	1	NRAM_IARUE NEURAMINIDASE (EC 3.2.	1.91e+01
41	47	47.5	512	1	AMY_BACLI ALPHA-AMYLASE PRECURSOR	1.91e+01
42	47	47.5	513	1	CATE_PSEAE CATALASE PRECURSOR (EC	1.91e+01
43	47	47.5	544	1	TCBP_MOUSE T-COMPLEX PROTEIN 1, E	1.91e+01
44	47	47.5	606	1	TERM_ADR12 DNA TERMINAL PROTEIN (1.91e+01
45	47	47.5	2278	1	FAB1_YEAST PROBABLE PHOSPHATIDYL	1.91e+01

ALIGNMENTS

RESULT	ID	STANDARD	PRT	999 AA.
AC	DSG3_HUMAN	P32926		
DT	01-OCT-1993 (REL. 27, CREATED)			
DT	01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	DESMOGLIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).			
GN	DSG3.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;			
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 92069753.			
RA	AMAGAI M., KLAUS-KOVUN V., STANLEY J.R.;			
RT	"Autoantibodies against a novel epithelial cadherin in pemphigus			
RT	vulgaris, a disease of cell adhesion."			
RL	CELL 67:869-877(1991).			
CC	-1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOGOME JUNCTIONS.			
CC	INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE			
CC	FILAMENT'S MEDIATING CELL-CELL ADHESION.			
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.			
CC	-1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, OESOPHAGUS AND			
CC	CARCINOMAS.			
CC	-1- DOMAIN: CALCINUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS			
CC	(POTENTIAL).			
CC	-1- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN			
CC	DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE			
CC	LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES			
CC	AGAINST DSG3.			
CC	-1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE			
CC	DESMOSOMAL SUBFAMILY.			
CC	-----			
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION			
CC	BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -			
CC	THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS			
CC	USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY			
CC	MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL			
CC	ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/			
CC	OR SEND AN EMAIL TO license@isb-sib.ch).			
CC	-----			
CC	EMBL; M76482; G190752; -.			
CC	PIR; A41088; ITHUG3.			
CC	MIM; 169615; -.			
CC	PROSITE; PS00237; CADHERIN; 3.			
CC	PFAM; PF00028; cadherin; 4.			
CC	HSSP; P09803; 1EDH.			

KW DNA-BINDING; NUCLEAR PROTEIN.
 FT ZN_FING 68 84 C2H2-TYPE (POTENTIAL).
 FT NON_TER 478 478
 SO SEQUENCE 478 AA; 54026 MW; 9D100D91 CRC32;
 Query Match 56.6%; Score 56; DB 1; Length 478;
 Best Local Similarity 53.8%; Pred. No. 2.09e-01;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 301 OAGWVSRKSG 313
 1 QSGTWRHSTGC 13

QY 1 QSGTWRHSTGC 13

RESULT 4
 ID FYN CHICK STANDARD; PRT: 533 AA.
 AC 005876; 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN (EC 2.7.1.112) (P59-FYN).
 GN FYN
 OS GALLUS GALLUS (CHICKEN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
 CC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WHITE LECHORN; TISSUE-MUSCLE;
 RA MEDLINE; 93205395.
 RA SUDOL M., GREULICH H., NEWMAN L., SARKAR A., SUKAGAWA J., YAMAMOTO T.;
 RT "A novel yes-related kinase, Yrk, is expressed at elevated levels in
 RT neural and hematopoietic tissues.";
 RL ONCOGENE 8:823-831(1993).
 CC -1- FUNCTION: IMPLICATED IN THE CONTROL OF CELL GROWTH.
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC -1- SUBUNIT: ASSOCIATES THROUGH ITS SH3 DOMAIN, TO THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC -1- TISSUE SPECIFICITY: THYMUS AND SPLEEN.
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
 CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).
 CC -----
 CC EMBL; X52841; G62862; -
 DR PIR; S36351; S36351.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS50001; SH2; 1.
 DR PROSITE; PS50002; SH3; 1.
 DR PFM; PF00017; SH2; 1.
 DR PFM; PF00018; SH3; 1.
 DR PFM; PF00069; PKINASE; 1.
 DR HSP; P06241; FYN.
 KW PROTO-ONCOGENE; TRANSFERASE; TYROSINE-PROTEIN KINASE; PHOSPHORYLATION;
 KW ATP-BINDING; MYRISTYLATION; SH3 DOMAIN; SH2 DOMAIN; PALMITATE;
 KW LIPID-BINDING;
 KW LIPID; 0 0 BY SIMILARITY.
 FT INT_MET 1 1 MYRISTATE (BY SIMILARITY).
 FT LIPID 2 2 PALMITATE (BY SIMILARITY).
 FT LIPID 5 5 PALMITATE (BY SIMILARITY).
 FT DOMAIN 81 142 SH3.
 FT DOMAIN 148 245 SH2.
 FT DOMAIN 267 520 PROTEIN KINASE.

FT MOD_RES 11 11 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 FT NP_BIND 273 281 ATP (BY SIMILARITY).
 FT BINDING 295 295 ATP (BY SIMILARITY).
 FT ACT_SITE 386 386 BY SIMILARITY.
 FT MOD_RES 416 416 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 527 527 PHOSPHORYLATION (BY SIMILARITY).
 SO SEQUENCE 533 AA; 60134 MW; 05970E9A CRC32;
 Query Match 55.6%; Score 55; DB 1; Length 533;
 Best Local Similarity 57.1%; Pred. No. 3.54e-01;
 Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 71 HTGLTRRGGTGYT 84
 1 QSGTWRHSTGC 14

QY 1 QSGTWRHSTGC 14

RESULT 5
 ID FYN MOUSE STANDARD; PRT: 533 AA.
 AC P39688; 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN (EC 2.7.1.112) (P59-FYN).
 GN FYN
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 CC RODENTIA; SCIUROGNATHI; MORIDAE; MURINAE; MUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE; 91175680.
 RA COOKE M.P., PERLMUTTER R.M.;
 RT "Expression of a novel form of the fyn proto-oncogene in
 RT hematopoietic cells.";
 RL NEW BIOL. 1:66-74(1989).
 CC [2]
 CC SEQUENCE FROM N.A.
 RA LEE C., KIM M.G., JEON S.H., PARK D.E., PARK S.D., SEONG R.H.;
 RT SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC [3]
 CC PALMITOYLATION.
 RP MEDLINE; 94019312.
 RA SHENOY-SCARIA A.M., TIMSON L.K., KWONG J., SHAW A.S., LUBLIN D.M.;
 RT "Palmitoylation of an amino-terminal cysteine motif of protein
 RT tyrosine kinases p56lck and p59fyn mediates interaction with
 RT G12coy1-phosphatidylinositol-anchored proteins.";
 RL MOL. CELL. BIOL. 13:6385-6392(1993).
 CC [4]
 CC PALMITOYLATION.
 RP MEDLINE; 95071286.
 RA KOEGL M., ZLATKINE P., LEY S.C., COURTHESIDE S.A., MAGEE A.I.;
 RT "Palmitoylation of multiple Src-family kinases at a homologous N-
 RT terminal motif.";
 RL BIOCHEM. J. 303:749-753(1994).
 CC [5]
 CC PALMITOYLATION.
 RP MEDLINE; 97345356.
 RA WOLVEN A., OKAMURA H., ROSENBLATT Y., RESH M.D.;
 RT "Palmitoylation of p59fyn is reversible and sufficient for plasma
 RT membrane association.";
 RL MOL. BIOL. CELL. 8:1159-1173(1997).
 CC [6]
 CC MYRISTOYLATION.
 RP MEDLINE; 96251668.
 RA GALEN L.K.T., LINDER M.E., SHAW A.S.;
 RT "Multiple features of the p59fyn src homology 4 domain define a motif
 RT for immune-receptor tyrosine-based activation motif (ITAM) binding
 RT and for plasma membrane localization.";
 RL J. CELL BIOL. 133:1007-1015(1996).
 CC -1- FUNCTION: IMPLICATED IN THE CONTROL OF CELL GROWTH.
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC -1- SUBUNIT: ASSOCIATES THROUGH ITS SH3 DOMAIN, TO THE P85 SUBUNIT OF

CC PHOSPHATIDYLINOSITOL 3-KINASE.
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL: M27266; G309241; -
DR EMBL: U70324; G1575677; -
DR PIR: A44991; A44991.
DR MGD: MGI:95602; FYN.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS50001; SH2; 1.
DR PROSITE: PS50002; SH3; 1.
DR PFAM: PF00017; SH2; 1.
DR PFAM: PF00018; SH3; 1.
DR PFAM: PF00069; PKINASE; 1.
DR HSP: P06241; FYN.
KW PROTO-ONCOGENE; TRANSFERASE; TYROSINE-PROTEIN KINASE; PHOSPHORYLATION;
KW ATP-BINDING; MYRISTYLATION; SH3 DOMAIN; SH2 DOMAIN; PALMITATE;
KW LIPOPROTEIN.
FT INIT_MET 0 0 BY SIMILARITY.
FT LIPID 1 1 MYRISTATE.
FT LIPID 2 2 PALMITATE.
FT LIPID 5 5 PALMITATE.
FT LIPID 81 142 SH3.
FT DOMAIN 148 245 SH2.
FT DOMAIN 267 520 PROTEIN KINASE.
FT NP_BIND 273 281 ATP (BY SIMILARITY).
FT BINDING 295 295 ATP (BY SIMILARITY).
FT ACT_SITE 386 386 BY SIMILARITY.
FT MOD_RES 416 416 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 533 AA; 59926 MW; R23566FA CRC32;
Db 71 HTGLTRTGTGT 84
QY 1 QSGTMRTRHSTGCT 14
Query Match 55.6%; Score 55; DB 1; Length 533;
Best Local Similarity 57.1%; Pred. No. 3.54e-01;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
RESULT 6
ID FYN_XENLA STANDARD; PRT; 536 AA.
AC P13406;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 37, LAST ANNOTATION UPDATE)
DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN (EC 2.7.1.112) (P59-FYN).
GN FYN.
OS XENOPUS LAEVIS (AFRICAN CLAMMED FROG).
CC EURARCTA; METAFOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
CC MESOBATRACHIA; PIPODEA; PIPIPAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90191723.
RA STEELE R.E., DENG J.C., GHOSH C.R., FERRO J.B.;
RT "Structure and expression of fyn genes in xenopus laevis.";
RL ONCOGENE 5:369-376(1990).
CC -1- FUNCTION: IMPLICATED IN THE CONTROL OF CELL GROWTH.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -1- SUBUNIT: ASSOCIATES THROUGH ITS SH3 DOMAIN, TO THE P85 SUBUNIT OF

CC PHOSPHATIDYLINOSITOL 3-KINASE.
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL: X52188; G64702; -
DR EMBL: M27502; G214163; -
DR PIR: A43806; A43806.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS50001; SH2; 1.
DR PROSITE: PS50002; SH3; 1.
DR PFAM: PF00017; SH2; 1.
DR PFAM: PF00018; SH3; 1.
DR PFAM: PF00069; PKINASE; 1.
DR HSP: P06241; FYN.
KW PROTO-ONCOGENE; TRANSFERASE; TYROSINE-PROTEIN KINASE; PHOSPHORYLATION;
KW ATP-BINDING; MYRISTYLATION; SH3 DOMAIN; SH2 DOMAIN; PALMITATE;
KW LIPOPROTEIN.
FT INIT_MET 0 0 BY SIMILARITY.
FT LIPID 1 1 MYRISTATE (BY SIMILARITY).
FT LIPID 2 2 PALMITATE (BY SIMILARITY).
FT LIPID 5 5 PALMITATE (BY SIMILARITY).
FT LIPID 81 142 SH3.
FT DOMAIN 148 245 SH2.
FT DOMAIN 270 523 PROTEIN KINASE.
FT MOD_RES 11 11 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT NP_BIND 276 284 ATP (BY SIMILARITY).
FT BINDING 298 298 ATP (BY SIMILARITY).
FT ACT_SITE 389 389 BY SIMILARITY.
FT MOD_RES 419 419 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 536 AA; 60715 MW; 87236822 CRC32;
Db 71 HTGLTRTGTGT 84
QY 1 QSGTMRTRHSTGCT 14
Query Match 55.6%; Score 55; DB 1; Length 536;
Best Local Similarity 57.1%; Pred. No. 3.54e-01;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
RESULT 7
ID FYN_XIPHE STANDARD; PRT; 536 AA.
AC P27446;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN (EC 2.7.1.112) (P59-FYN).
GN FYN.
OS XIPHOPHUS HELLERI.
CC EURARCTA; METAFOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
CC TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; ATHERINOMORPHA.
CC CIPRINODONTIFORMES; CIPRINODONTIOIDEI; POECILIIDAE; XIPHOPHORUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-RIO LANCETILLA.
RC MEDLINE: 91187435.
RA HANNIG G., OTTILE S., SCHARF M.;
RT "Conservation of structure and expression of the c-yes and fyn genes
RT in lower vertebrates";
RL ONCOGENE 6:361-369(1991).
CC -1- FUNCTION: IMPLICATED IN THE CONTROL OF CELL GROWTH.

CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC -1- SUBUNIT: ASSOCIATES THROUGH ITS SH3 DOMAIN, TO THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
 CC DOMAIN, BELONGS TO THE SRC SUBFAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X54971; G6482; -
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS50001; SH2; 1.
 DR PROSITE; PS50002; SH3; 1.
 DR PRAM; PF00017; SH2; 1.
 DR PRAM; PF00018; SH3; 1.
 DR PRAM; PF00069; PKINASE; 1.
 DR HSSP; P06241; 1AOT.
 KM PROTO-ONCOGENE; TRANSFERASE; TYROSINE-PROTEIN KINASE; PHOSPHORYLATION;
 KM AMP-BINDING; MYRISTYLATION; SH3 DOMAIN; SH2 DOMAIN; PALMITATE;
 KM LIPOPROTEIN.
 KW INT_MET 0 BY SIMILARITY.
 FT LIPID 1 MYRISTATE (BY SIMILARITY).
 FT LIPID 2 PALMITATE (BY SIMILARITY).
 FT LIPID 5 PALMITATE (BY SIMILARITY).
 FT DOMAIN 81 142 SH2.
 FT DOMAIN 148 245 SH3.
 FT DOMAIN 270 523 PROTEIN KINASE.
 FT MOD_RES 11 11 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 FT NP_BIND 276 284 ATP (BY SIMILARITY).
 FT BINDING 298 298 ATP (BY SIMILARITY).
 FT ACT_SITE 389 389 BY SIMILARITY.
 FT MOD_RES 419 419 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 530 530 PHOSPHORYLATION (BY SIMILARITY).
 SQ SEQUENCE 536 AA; 60316 MW; BA03DB12 CRC32;
 Query Match 55.6%; Score 55; DB 1; Length 536;
 Best Local Similarity 57.1%; Pred. No. 3.54e-01;
 Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 71 HTGTLRTGSGTGT 84
 ::::|||||
 Oy 1 QSGTMRTRHSGTGT 14

RESULT 8
 ID FYN_HUMAN STANDARD; PRT; 536 AA.
 AC P06241;
 DT 01-JAN-1988 (REL. 06, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN (EC 2.7.1.112) (P59-FYN)
 DE (SYN) (SLK).
 GN FYN.
 OS HOMO SAPIENS (HUMAN).
 OS EURARCTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE; 86287278.
 RA SEMBA K., NISHIZAWA M., MIYAJIMA N., YOSHIDA M.C., SUKAGAWA J.,
 RA YAMANASHI Y., SASAKI M., YAMAMOTO T., TOROSHIMA K.;
 RA "Yas-related protooncogene, syn, belongs to the protein-tyrosine
 RT kinase family.";

RL PROC. NATL. ACAD. SCI. U.S.A. 83:5459-5463(1986).
 RN [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE; 87089775.
 RA KAWAKAMI T., PENNINGTON C.Y., ROBBINS K.C.;
 RT "Isolation and oncogenic potential of a novel human src-like gene."
 RL MOL. CELL. BIOL. 6:4195-4201(1986).
 RN [3]
 RN MYRISTOYLATION, AND PHOSPHORYLATION AT TYR-530.
 RX MEDLINE; 91016431.
 RA PETERS D.J., MCGREW B.R., PERRON D.C., LIPIAK L.M., LAUDANO A.P.;
 RT "In vivo phosphorylation and membrane association of the fyn proto-
 RT oncogene product in IM-9 human lymphoblasts."
 RL ONCOGENE 5:1313-1319(1990).
 RN [4]
 RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF SH3 DOMAIN.
 RX MEDLINE; 93327750.
 RA NOBLE M.E.M., MUSACCHIO A., SARASTE M., COURTNEIDGE S.A.,
 RA WIRENGA R.K.;
 RT "Crystal structure of the SH3 domain in human Fyn: comparison of the
 RT three-dimensional structures of SH3 domains in tyrosine kinases and
 RT spectrin."
 RL EMBO J. 12:2617-2624(1993).
 RN [5]
 RN X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 80-141.
 RX MEDLINE; 95393198.
 RA MUSACCHIO A., SARASTE M., WILMANS M.;
 RT "High-resolution crystal structures of tyrosine kinase SH3 domains
 RT complexed with proline-rich peptides."
 RL NAT. STRUCT. BIOL. 1:546-551(1994).
 RN [6]
 RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 84-140 IN COMPLEX WITH NEF.
 RX MEDLINE; 96279837.
 RA LEE C.H., SAKSELA K., MIRZA U.A., CHAIT B.T., KURIYAN J.;
 RT "Crystal structure of the conserved core of HIV-1 Nef complexed with
 RT a Src family SH3 domain."
 RL CELL 85:931-942(1996).
 RN [7]
 RN STRUCTURE BY NMR OF SH3 DOMAIN.
 RP MEDLINE; 96399716.
 RA MORTON C.J., PUGH D.J.R., BROWN E.L.J., KAHMANN J.D., RENZONI D.A.C.,
 RA CAMPBELL I.D.;
 RT "Solution structure and peptide binding of the SH3 domain from human
 RT Fyn."
 RL STRUCTURE 4:705-714(1996).
 RN [8]
 RN STRUCTURE BY NMR.
 RP MEDLINE; 97121261.
 RA RENZONI D.A., PUGH D.J., SILIGARDI G., DAS P., MORTON C.J., ROSSI C.,
 RA WATERFIELD M.D., CAMPBELL I.D., LADBURY J.E.;
 RT "Structural and thermodynamic characterization of the interaction of
 RT the SH3 domain from Fyn with the proline-rich binding site on the p85
 RT subunit of PI3-kinase."
 RL BIOCHEMISTRY 35:15646-15653(1996).
 RN [9]
 RN STRUCTURE BY NMR OF SH2 DOMAIN.
 RP MEDLINE; 98035454.
 RA MUIHERN T.D., SHAW G.L., MORTON C.J., DAY A.J., CAMPBELL I.D.;
 RT "The SH2 domain from the tyrosine kinase Fyn in complex with a
 RT phosphotyrosyl peptide reveals insights into domain stability and
 RT binding specificity."
 RL STRUCTURE 5:1313-1323(1997).
 RN [10]
 RN BINDING OF SH3 DOMAIN TO PI 3-KINASE.
 RP MEDLINE; 93348274.
 RA PRASAD K.V., JANSSEN O., KAPILLER R., RAAB M., CANTLEY L.C.,
 RA RUDD C.E.;
 RT "src-homology 3 domain of protein kinase p59fyn mediates binding to
 RT phosphatidylinositol 3-kinase in T cells."
 RL PROC. NATL. ACAD. SCI. U.S.A. 90:7366-7370(1993).
 CC -1- FUNCTION: IMPLICATED IN THE CONTROL OF CELL GROWTH.
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
 CC PROTEIN TYROSINE PHOSPHATE.

CC -1- SUBUNIT: ASSOCIATES THROUGH ITS SH3 DOMAIN, TO THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE.
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M14333; G181172; -;
DR EMBL; M14676; G338228; -;
DR PIR; A24314; TVH05Y;
DR PIR; A25389; TVH05Y;
DR PDB; 1SHF; 31-OCT-93.
DR PDB; 1FVN; 08-NOV-96.
DR PDB; 1NFG; 08-NOV-96.
DR PDB; 1NFG; 08-NOV-96.
DR PDB; 1EON; 11-JAN-97.
DR PDB; 1AON; 25-FEB-98.
DR PDB; 1AOT; 14-JAN-98.
DR PDB; 1AOU; 14-JAN-98.
DR PDB; 1AZG; 25-FEB-98.
DR MIM; 137025; -;
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PFM; PF00017; SH2; 1.
DR PFM; PF00018; SH3; 1.
DR PFM; PF00069; PKINASE; 1.
DR PROTO-ONCOGENE; TRANSFERASE; TYROSINE-PROTEIN KINASE; PHOSPHORYLATION;
KW ATP-BINDING; MYRISTYLATION; SH3 DOMAIN; SH2 DOMAIN; PALMITATE;
KW LIPOPROTEIN; 3D-STRUCTURE.
KW INIT_MET 0
FT LIPID 1 MYRISTATE.
FT LIPID 2 PALMITATE (BY SIMILARITY).
FT LIPID 5 PALMITATE (BY SIMILARITY).
FT DOMAIN 81 SH2.
FT DOMAIN 148 SH3.
FT DOMAIN 245 SH2.
FT DOMAIN 270 SH3.
FT MOD_RES 11 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT NP_BIND 276 ATP (BY SIMILARITY).
FT BINDING 298 ATP (BY SIMILARITY).
FT ACT_SITE 389 BY SIMILARITY.
FT MOD_RES 419 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 530 PHOSPHORYLATION.
FT CONFLICT 183 A -> S (IN REF. 2).
FT CONFLICT 436 A -> R (IN REF. 2).
FT STRAND 85
FT STRAND 92
FT STRAND 99
FT STRAND 102
FT TURN 104
FT STRAND 107
FT STRAND 112
FT STRAND 118
FT TURN 124
FT STRAND 129
FT STRAND 133
FT HELIX 134
FT STRAND 137
SQ SEQUENCE 536 AA; 60630 MW; 01B18DD0 CRC32;

Db 71 HTGLTRGTGTGT 84

Query Match 55.6%; Score 55; DB 1; Length 536;
Best Local Similarity 57.1%; Pred. No. 3.34e-01;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY :||:||||:|
1 QSGTMRTRHSTGT 14
RESULT 9
ID TCEP_HUMAN STANDARD; PRT; 543 AA.
AC Q99832; O14871;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE T-COMPLEX PROTEIN 1, ETA SUBUNIT (TCEP-1-ETA) (CCT-ETA) (HIV-1 NEF
DE INTERACTING PROTEIN).
GN CCT7 OR CCT8 OR NIP7-1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA MON K.-A., REED S.L.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE OF 13-424 FROM N.A.
RA FUKUSHI M., KIMURA T., YAMAMOTO N.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: MOLECULAR CHAPERONE; ASSIST THE FOLDING OF PROTEINS UPON
CC ATP HYDROLYSIS. KNOWN TO PLAY A ROLE, IN VITRO, IN THE FOLDING OF
CC ACTIN AND TUBULIN (BY SIMILARITY).
CC -1- SUBUNIT: HETERO-OLIGOMERIC COMPLEX OF ABOUT 850 TO 900 KD THAT
CC FORMS TWO STACKED RINGS, 12 TO 16 NM IN DIAMETER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TCP-1 CHAPERONIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF026292; G2559010; -;
DR EMBL; U83843; G1800303; -;
DR PROSITE; PS00750; TCP1_1; 1.
DR PROSITE; PS00751; TCP1_2; 1.
DR PROSITE; PS00995; TCP1_3; 1.
DR PFM; PF00118; cpn60_TCP1; 1.
DR HSSP; P48424; IASX.
KW CHAPERONE; ATP-BINDING; MULTIGENE FAMILY.
FT CONFLICT 282 HH -> RO (IN REF. 2).
FT CONFLICT 293 L -> P (IN REF. 2).
FT CONFLICT 336 C -> L (IN REF. 2).
FT CONFLICT 364 C -> L (IN REF. 2).
FT CONFLICT 374 LRG -> SPC (IN REF. 2).
FT CONFLICT 407 A -> P (IN REF. 2).
FT CONFLICT 411 A -> P (IN REF. 2).
SQ SEQUENCE 543 AA; 59366 MW; A1023F61 CRC32;

Query Match 52.5%; Score 52; DB 1; Length 543;
Best Local Similarity 60.0%; Pred. No. 1.67e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 464 LRARHAGCT 473
OY 5 MTRHSTGT 14

RESULT 10
ID PABP_STRLL STANDARD; PRT; 475 AA.
AC P27630;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PARA-AMINOBENZOATE SYNTHASE COMPONENT I (EC 4.1.1.3.-) (ADC SYNTHASE).

GN PABB.
OS STREPTOMYCES LIVIDANS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=66 / 1326;
RX MEDLINE: 93231526.
RA ARHIN F.F., VINING L.C.;
RT "Organization of the genes encoding p-aminobenzoic acid synthetase
from Streptomyces lividans 1326.";
RL GENE 126:128-133(1993).
CC -1- FUNCTION: CATALYZES THE BIOSYNTHESIS OF 4-AMINO-4-DEOXYCHORISMATE
(ADC) FROM CHORISMATE AND GLUTAMINE.
CC -1- PATHWAY: FOLATE BIOSYNTHESIS PATHWAY. FIRST STEP IN THE
BIOSYNTHESIS OF P-AMINOBENZOATE (PABA).
CC -1- SUBUNIT: CONSISTS OF TWO NONIDENTICAL CHAINS: COMPONENT I
CATALYZES THE FORMATION OF ADC BY BINDING CHORISMATE AND AMMONIA;
CC COMPONENT II PROVIDES THE GLUTAMINE AMIDOTRANSFERASE ACTIVITY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M64859; GI533394; -
DR PIR: JN0578; JN0578.
DR PRAM: PF00425; chorismate_bind; 1.
KW LIASE; FOLATE BIOSYNTHESIS.
SQ SEQUENCE 475 AA; 52042 MW; 9A70C557 CRC32;

Query Match 51.5%; Score 51; DB 1; Length 475;
Best Local Similarity 42.9%; Pred. No. 2.76e+00;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 281 AGTLGRPRAGDD 294
QY 2 SGTMRTRHSTGCTN 15

RESULT 11
ID GLC1-SOYBN STANDARD; PRT; 481 AA.
AC P11828;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GLYCININ G3 PRECURSOR [CONTAINS: GLYCININ A SUBUNIT; GLYCININ B
SUBUNIT].
GN GY3.
OS GLYCINE MAX (SOYBEAN).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMERYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC FABALES; FABACEAE; PAPILIONOIDEAE; GLYCINE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. DARE; TISSUE=LEAF;
RX MEDLINE: 89296500.
RA CHO T.-J., NIELSEN N.C.;
RI "The glycinin GY3 gene from soybean.";
RL NUCLEIC ACIDS RES. 17:4388-4388(1989).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE: 92393391.
RA NIELSEN N.C., DICKINSON C.D., CHO T.-J., THANH V.H., SCALLON B.J.,
RA FISCHER R.L., SIMS T.L., DREWS G.N., GOLDBERG R.B.;
RT "Characterization of the glycinin gene family in soybean.";
RL PLANT CELL 1:313-328(1989).
CC -1- FUNCTION: GLYCININ IS THE MAJOR SEED STORAGE PROTEIN OF SOYBEAN.
CC -1- SUBUNIT: HEXAMER; EACH SUBUNIT IS COMPOSED OF AN ACIDIC AND A
BASIC CHAIN DERIVED FROM A SINGLE PRECURSOR AND LINKED BY A

CC DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE 11S SEED STORAGE PROTEINS (GLOBULINS)
FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X15123; G18639; -
DR PIR: S04605; S04605.
DR PROSITE: PS00305; 11S_SEED_STORAGE; 1.
DR PRAM: PF00190; Seedstore_11s; 1.
KW SEED STORAGE PROTEIN; MULTIGENE FAMILY; SIGNAL.
FT SIGNAL 1 21
FT CHAIN 22 296 A SUBUNIT.
FT CHAIN 297 476 B SUBUNIT.
FT PROPEP 477 481
FT DISULFID 107 303 INTERCHAIN (BY SIMILARITY).
SQ SEQUENCE 481 AA; 54241 MW; B1A9BDFO CRC32;

Query Match 51.5%; Score 51; DB 1; Length 481;
Best Local Similarity 58.3%; Pred. No. 2.76e+00;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 304 TMLRLRHNGQTS 315
QY 4 TMRTRHSTGCTN 15

RESULT 12
ID GLC1-SOYBN STANDARD; PRT; 495 AA.
AC P04776;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GLYCININ G1 PRECURSOR [CONTAINS: GLYCININ A1A SUBUNIT; GLYCININ B
SUBUNIT].
GN GY1.
OS GLYCINE MAX (SOYBEAN).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMERYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC FABALES; FABACEAE; PAPILIONOIDEAE; GLYCINE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. DARE; TISSUE=LEAF;
RX MEDLINE: 92393391.
RA NIELSEN N.C., DICKINSON C.D., CHO T.-J., THANH V.H., SCALLON B.J.,
RA FISCHER R.L., SIMS T.L., DREWS G.N., GOLDBERG R.B.;
RT "Characterization of the glycinin gene family in soybean.";
RL PLANT CELL 1:313-328(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. DARE; TISSUE=LEAF;
RX MEDLINE: 89296498.
RA SIMS T.L., GOLDBERG R.B.;
RT "The glycinin GY1 gene from soybean.";
RL NUCLEIC ACIDS RES. 17:4386-4386(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. BONMINORI;
RX MEDLINE: 86041867.
RA NEGRO T., MOMMA T., FUKAZAWA C.;
RT "A cDNA clone encoding a glycinin Ala subunit precursor of soybean.";
RL NUCLEIC ACIDS RES. 13:6719-6731(1985).
RN [4]
RP SEQUENCE FROM N.A.
RA UTSUMI S., KOHNO M., MORI T., KITO M.;
RT "An alternate cDNA encoding glycinin Ala Bx subunit.";
RL J. AGRIC. FOOD CHEM. 35:210-214(1987).

CC -1- FUNCTION: GLYCININ IS THE MAJOR SEED STORAGE PROTEIN OF SOYBEAN.
 CC -1- SUBUNIT: HEXAMER; EACH SUBUNIT IS COMPOSED OF AN ACIDIC AND A
 CC BASIC CHAIN DERIVED FROM A SINGLE PRECURSOR AND LINKED BY A
 CC DISULFIDE BOND.
 CC -1- PTM: THE PRECURSOR IS POST-TRANSLATIONALLY PROCESSED TO FORM A
 CC COVALENTLY LINKED A1A-BX SUBUNIT COMPLEX.
 CC -1- SIMILARITY: BELONGS TO THE 11S SEED STORAGE PROTEINS (GLOBULINS)
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between, the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M3686; G169973; -
 CC EMBL: X15121; G18635; -
 CC EMBL: X02985; G18615; -
 CC PIR: A23497; EMSY2.
 CC PIR: JS0015; JS0015.
 CC PIR: S04603; S04603.
 CC PIR: S10851; S10851.
 CC PROSITE: PS00305; 11S_SEED_STORAGE; 1.
 CC PRAM: PF00190; Seedstore_11s; 1.
 CC SEED STORAGE PROTEIN; MULTIGENE FAMILY; SIGNAL.
 CC SIGNAL
 CC FT CHAIN 1 19
 CC FT PROPEP 20 306 A1A SUBUNIT.
 CC FT CHAIN 307 310 BX SUBUNIT.
 CC FT CHAIN 311 490
 CC FT PROPEP 491 495
 CC FT DISULFID 107 317 INTERCHAIN (BY SIMILARITY).
 CC FT CONFLICT 42 42 D -> G (IN REF. 3).
 CC FT CONFLICT 108 108 P -> S (IN REF. 3).
 CC FT CONFLICT 136 136 F -> S (IN REF. 3).
 CC FT CONFLICT 360 360 E -> G (IN REF. 3).
 CC SEQUENCE 495 AA; 55706 MW; 32A73948 CRC32;
 CC
 CC Query Match 51.5%; Score 51; DB 1; Length 495;
 CC Best Local Similarity 58.3%; Pred. No. 2.76e+00;
 CC Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 CC
 CC Db 318 TMRLRHIGQTS 329
 CC |||::|:
 CC 4 TMRLRHSTGGIN 15
 CC
 CC RESULT 13
 CC ID M45D_SCHTA STANDARD; PRT; 120 AA.
 CC AC P13411;
 CC DT 01-JAN-1990 (REL. 13, CREATED)
 CC RT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
 CC NT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 CC DE IMMUNOGENIC MIRACIDIAL ANTIGEN 5D (FRAGMENT).
 CC GN 5D.
 CC OS SCHISTOSOMA JAPONICUM (BLOOD FLUKE).
 CC UC EUKARYOTA; METAZOA; PLATHELMINTHES; TREMATODA; DIGENEA; STRIGEIDIDA;
 CC OC SCHISTOSOMATOIDEA; SCHISTOSOMATIDAE; SCHISTOSOMA.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN-PHILIPPINE;
 CC RX MEDLINE: 89261924.
 CC PA SCALION B.J., BOGITSH B.J., CARTER C.E.;
 CC RT Characterisation of a large gene family in Schistosoma japonicum
 CC that encodes an immunogenic miracidial antigen.*;
 CC RL MOL. BIOCHEM. PARASITOL. 33:105-112(1989).
 CC CC -1- DEVELOPMENTAL STAGE: MIRACIDIA.
 CC CC -1- SIMILARITY: NEARLY IDENTICAL TO IMMUNOGENIC MIRACIDIAL
 CC ANTIGENS 81, 81 AND 8C.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M25214; G552238; -
 CC DR ANTIGEN: MULTIGENE FAMILY.
 CC FT NON_TER 1 1
 CC SEQUENCE 120 AA; 13541 MW; C5BE9039 CRC32;
 CC
 CC Query Match 50.5%; Score 50; DB 1; Length 120;
 CC Best Local Similarity 57.1%; Pred. No. 4.53e+00;
 CC Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
 CC
 CC Db 77 QGGYQNRHAYGGT 90
 CC |||::|:
 CC 1 QSGTMRHSTGTT 14
 CC
 CC RESULT 14
 CC ID AAC8_STRPR STANDARD; PRT; 286 AA.
 CC AC P29809;
 CC DT 01-APR-1993 (REL. 25, CREATED)
 CC DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
 CC DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
 CC DE AMINOGLYCOSIDE N3'-ACETYLTTRANSFERASE VIII (EC 2.3.1.81) (ACC(3)-VIII)
 CC DE (AMINOGLYCOSIDE 3-N-ACETYLTTRANSFERASE TYPE VIII).
 CC GN AAC8.
 CC OS STREPTOMYCES FRADIAE.
 CC OC BACTERIA; FRADICTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
 CC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE: 91285425.
 CC RA SALAZAR D., PEREZ-GONZALEZ J.A., PIEPERSBERG W., DAVIES J.;
 CC RT Characterisation of aminoglycoside acetyltransferase encoding genes
 CC of neomycin-producing Micromonospora chalybea and Streptomyces
 CC fradiae.*;
 CC RL GENE 101:143-148(1991).
 CC CC -1- FUNCTION: RESISTANCE TO NEOMYCIN.
 CC CC -1- CATALYTIC ACTIVITY: ACETYL-COA + A 2-DEOXYSTREPTAMINE ANTI-BIOTIC
 CC -COA + N3'-ACETYL-2-DEOXYSTREPTAMINE ANTI-BIOTIC.
 CC CC -1- SIMILARITY: TO OTHER AAC(3) PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M55426; G295164; -
 CC DR PIR: JEO417; JEO417.
 CC KW ANTI-BIOTIC RESISTANCE; TRANSFERASE; ACYLTRANSFERASE.
 CC SEQUENCE 286 AA; 30455 MW; 44C39955 CRC32;
 CC
 CC Query Match 50.5%; Score 50; DB 1; Length 286;
 CC Best Local Similarity 38.5%; Pred. No. 4.53e+00;
 CC Matches 5; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC Db 120 QGAVRSRHPDAS 132
 CC |||::|:
 CC 1 QSGTMRHSTGTT 13
 CC
 CC RESULT 15
 CC ID HEMA_P13B STANDARD; PRT; 572 AA.
 CC AC P06167;
 CC DT 01-JAN-1988 (REL. 06, CREATED)
 CC DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
 CC DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)

```

DE HEMAGGLUTININ-NEURAMINIDASE (EC 3.2.1.18).
GN HN.
OS BOVINE PARAINFLUENZA 3 VIRUS.
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; PARAMYXOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-910N;
RX MEDLINE; 87174819.
RA SUGU S., SAKAI Y., SHIODA T., SHIBUYA H.;
RT "Nucleotide sequence of the bovine parainfluenza 3 virus genome: the
RL genes of the F and HN glycoproteins";
RL NUCLEIC ACIDS RES. 15:2945-2958(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS
CC TO CELL RECEPTORS AND FOR INITIATING INFECTION. NEURAMINIDASE
CC ACTIVITY HELPS THE EFFICIENT SPREAD OF THE VIRUS BY DISSOCIATING
CC THE MATURE VIRIONS FROM THE NEURAMINIC ACID CONTAINING
CC GLYCOPROTEINS.
CC -1- SUBCELLULAR LOCATION: EXTERNAL, ANCHORED TO THE ENVELOPE BY ITS
CC N-TERMINAL HYDROPHOBIC SEQUENCE.
CC -1- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES HEMAGGLUTININ-
CC NEURAMINIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y00114; G60897; -
DR EMBL; D84095; G1255658; -
DR PIR; B27218; HNNZB3.
DR PFAM; PF00423; HN; 1.
KW HYDROLASE; HEMAGGLUTININ; ENVELOPE PROTEIN; GLYCOPROTEIN;
KW TRANSMEMBRANE.
FT DOMAIN 1 30 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 31 53 POTENTIAL.
FT DOMAIN 54 572 EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 308 308 POTENTIAL.
FT CARBOHYD 351 351 POTENTIAL.
FT CARBOHYD 448 448 POTENTIAL.
FT CARBOHYD 523 523 POTENTIAL.
FT CARBOHYD 570 570 POTENTIAL.
SQ SEQUENCE 572 AA; 64590 MM; 876CFEE6 CRC32;

Query Match 50.5%; Score 50; DB 1; Length 572;
Best Local Similarity 61.5%; Pred. No. 4.53e+00;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

Db 18 GTTRDRHSSKATN 30
 ||| |||:::|
 OY 3 GTMRTRHSTGGTN 15

Search completed: Fri Jun 11 17:52:31 1999
 Job time : 9 secs.

 MIPSE

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit,
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mpsrch_p protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Jun 11 17:54:53 1999; Maspar time 6.10 Seconds

Tabular output not generated. 134.225 Million cell updates/sec

Title: >US-08-991-628-7
 Description: (1-15) from US08991628.pep
 Perfect Score: 99
 Sequence: 1 QSGTMRTRHSTGTG 15

Scoring table: PAM 150
 Gap 15

Searched: 179066 segs. 54579741 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: sptrembl9
 1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
 5:sp_invertebrate 6:sp_mammal 7:sp_mmc 8:sp_organelle
 9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
 13:sp_vertebrate 14:sp_virus

Statistics: Mean 24.193; Variance 26.150; scale 0.925

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	87.9	993	11	Q35902	DESMOGLEIN 3 (FRAGMENT	9.09e-10
2	82.6	421	2	Q08590	POTATIVE SECRETED PROT	8.35e-03
3	57.6	1597	5	Q61346	SMALL OPTIC LOBES.	1.48e-01
4	55.6	534	4	Q16248	P59FYN.	4.50e-01
5	55.6	537	11	Q62844	PROTO-ONCOGENE FYN.	4.50e-01
6	54.5	776	11	Q08511	DNA CYTOSINE-5 METHYL	7.76e-01
7	54.5	839	11	Q08510	DNA CYTOSINE-5 METHYL	7.76e-01
8	54.5	859	11	Q08509	DNA CYTOSINE-5 METHYL	7.76e-01
9	53.5	160	9	Q08162	COMPLETE GENOME.	1.33e+00
10	53.5	160	9	Q08162	COMPLETE GENOME.	1.33e+00
11	53.5	160	9	Q08162	COMPLETE GENOME.	1.33e+00
12	53.5	161	9	Q08162	COMPLETE GENOME.	1.33e+00
13	53.5	161	9	Q08162	COMPLETE GENOME.	1.33e+00
14	53.5	161	9	Q08162	COMPLETE GENOME.	1.33e+00
15	53.5	161	9	Q08162	COMPLETE GENOME.	1.33e+00
16	52.5	136	2	Q051792	HYPOHETICAL 14.7 KD P	2.26e+00
17	52.5	136	2	Q051792	HYPOHETICAL 14.7 KD P	2.26e+00
18	52.5	1042	5	P90974	ADAM-1 PREPROTEIN PR	2.26e+00
19	52.5	1921	5	Q01349	KINESIN-73.	2.26e+00
20	51.5	69	14	P90363	GLYCOPROTEIN B (FRAGME	3.82e+00

21	51	51.5	69	14	P90292	GLYCOPROTEIN B (FRAGME	3.82e+00
22	51	51.5	83	14	Q68570	GLYCOPROTEIN B (FRAGME <td>3.82e+00</td>	3.82e+00
23	51	51.5	169	14	P90393	GLYCOPROTEIN B (FRAGME <td>3.82e+00</td>	3.82e+00
24	51	51.5	170	14	P90392	GLYCOPROTEIN B (FRAGME <td>3.82e+00</td>	3.82e+00
25	51	51.5	171	14	P89538	GLYCOPROTEIN B (FRAGME <td>3.82e+00</td>	3.82e+00
26	51	51.5	172	14	P90394	GLYCOPROTEIN B (FRAGME <td>3.82e+00</td>	3.82e+00
27	51	51.5	178	14	P69174	GLYCOPROTEIN B (FRAGME <td>3.82e+00</td>	3.82e+00
28	51	51.5	797	5	Q44536	COSMID H22D07.	3.82e+00
29	51	51.5	905	14	Q69173	GLYCOPROTEIN B (FRAGME <td>3.82e+00</td>	3.82e+00
30	50	50.5	74	14	Q11346	CLONE H1-41 HOMOLOG OF <td>6.40e+00</td>	6.40e+00
31	50	50.5	275	2	Q44378	VIRULENCE PROTEIN.	6.40e+00
32	50	50.5	396	5	Q22250	T0658.3 PROTEIN.	6.40e+00
33	50	50.5	505	2	Q53827	CEPHALOSPORIN C HYDROX <td>6.40e+00</td>	6.40e+00
34	50	50.5	572	14	Q65689	HEMAGGLUTININ-NEURAMIN <td>6.40e+00</td>	6.40e+00
35	49	49.5	140	2	Q45261	NOD L.	1.06e+01
36	49	49.5	341	2	Q07863	RPKA.	1.06e+01
37	49	49.5	368	11	Q61642	HISOCOMPATIBILITY 2,	1.06e+01
38	49	49.5	368	7	Q61895	PRECURSOR.	1.06e+01
39	49	49.5	584	5	Q25289	MAJOR SURFACE GLYCOPRO	1.06e+01
40	49	49.5	1138	4	Q43166	KIAA0440 (FRAGMENT).	1.06e+01
41	49	49.5	1502	3	Q60081	HYPOHETICAL 171.2 KD	1.06e+01
42	49	49.5	1822	11	Q35412	SPA-1 LIKE PROTEIN P12	1.06e+01
43	49	49.5	1907	4	Q15017	KIAA0299 (FRAGMENT).	1.06e+01
44	49	49.5	5071	5	P91905	RYANODINE RECEPTOR.	1.06e+01
45	49	49.5	5107	5	Q94279	PARTIAL CDS.	1.06e+01

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	993 AA.
ID	Q35902			
AC	Q35902			
DT	01-JAN-1998	(TREMBLREL. 05, CREATED)		
DT	01-JAN-1998	(TREMBLREL. 05, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)		
DE	DESMOGLEIN 3 (FRAGMENT).			
GN	DSG3.			
OS	MUS MUSCULUS (MOUSE).			
OC	EUCARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;			
NC	SCUROGNATHI; MORIDAE; MORINAE; MUS.			
RC	SEQUENCE FROM N.A.			
RA	ISHIKAWA H., Li K., UETTO J.;			
RL	SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.			
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).			
DR	EMBL: U86016; G2290200; -			
DR	PROSITE: P800232; CADHERIN. 2.			
DR	PFAW; PF00028; Cadherin; 4.			
KW	CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; CALCIUM-BINDING; REPEAT.			
FT	NON_TER 993			
FT	SEQUENCE 993 AA; 107888 MW; 881794BD-CRC32;			
Query Match	87.9%;	Score 87;	DB 11;	Length 993;
Best Local Similarity	100.0%;	Pred. 9.09e-10;		
Matches	13;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;
DB	743	SGTMRTRHSTGT 755		
QY	2	SGTMRTRHSTGT 14		
RESULT	2	PRELIMINARY:	PRT:	421 AA.
ID	Q08590			
AC	Q08590			
DT	01-NOV-1998	(TREMBLREL. 08, CREATED)		
DT	01-NOV-1998	(TREMBLREL. 08, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)		
DE	POTATIVE SECRETED PROTEIN.			
GN	SC2H4.07C.			
OS	STREPTOMYCES COELICOLOR.			
OC	BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;			
OC	ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.			

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA OLIVER K., HARRIS D.;
RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA REDENBACH M., KIESER H.M., DENAPAITE D., EICHNER A., CULLUM J.,
RA KINASHI H., HOPWOOD D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL EMBL: AL031514; E1319225;
DR SEQUENCE 421 AA; 44174 MW; 7B444909 CRC32;
SQ
Query Match 62.6%; Score 62; DB 2; Length 421;
Best Local Similarity 69.2%; Pred. No. 8.35e-03;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
DB 48 SGMTRHSTGCT 60
QY 2 SGMTRHSTGCT 14

```

```

QY [1]
QY 2 SGMTRHSTGCT 14
RESULT 4
ID Q16248; PRELIMINARY; PRT; 534 AA.
AC Q16248;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE P59FTN.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
CA CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE; 95123064.
RA RIGLEY K., SLOCOMBE P., PROUDFOOT K., WAHID S., MANDAIR K.,
RA BBRINGTON C.;
RT "Human p59fyn(1) regulates OKT3-induced calcium influx by a mechanism
RT distinct from p12 hydrolysis in Jurkat T cells.";
RL J. IMMUNOL. 154:1136-1145(1995).
DR EMBL: S74774; G802051.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; PKINASE; 1.
DR SEQUENCE 534 AA; 60226 MW; D3941652 CRC32;
SQ
Query Match 55.6%; Score 55; DB 4; Length 534;
Best Local Similarity 57.1%; Pred. No. 4.50e-01;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
DB 72 HTGLTRHSTGCT 85
QY 1 QSGTMRHSTGCT 14
RESULT 5
ID Q62844; PRELIMINARY; PRT; 537 AA.
AC Q62844;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE P59FTN.
OS RATRUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
CA SCIUROGNATHI; MURIDAE; MURINAE; RATRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SHSP STROKE-PRONE SPONTANEOUSLY HYPERTENSIVE;
RC TISSUE-WHOLE BRAIN;
RA NEMOTO K., SEKIMOTO M., KAGEYAMA H., FUKAMACHI K., NEMOTO F.,
RA UDEYAMA T., SENBA E., TOMITA I.;
RL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U33365; G1101768.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; PKINASE; 1.
DR SEQUENCE 537 AA; 60701 MW; D0A1DD46 CRC32;
SQ
Query Match 55.6%; Score 55; DB 11; Length 537;
Best Local Similarity 57.1%; Pred. No. 4.50e-01;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
DB 72 HTGLTRHSTGCT 85
QY 1 QSGTMRHSTGCT 14

```

```

RESULT 6
ID 008511 PRELIMINARY; PRT; 776 AA.
AC 008511;
DB 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
GN DNA CYTOSINE-5 METHYLTRANSFERASE 3B3.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
SCUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98324766.
RA OKANO M., XIE S., LI E.;
RT "Cloning and characterization of a family of novel mammalian DNA
(cytosine-5) methyltransferases."
RL NAT. GENET. 19:219-220(1998).
DR EMBL; AF068628; G3327984; -.
KW TRANSFERASE; METHYLTRANSFERASE.
SQ SEQUENCE 776 AA; 87590 MW; 712A1E2 CRC32;

Query Match 54.5%; Score 54; DB 11; Length 776;
Best Local Similarity 46.7%; Pred. No. 7.76e-01;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 115 ESPAVTRHSGTSS 129
OY 1 QSGTMRTRHSGTGN 15

RESULT 7
ID 008510 PRELIMINARY; PRT; 839 AA.
AC 008510;
DB 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
GN DNA CYTOSINE-5 METHYLTRANSFERASE 3B2.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
SCUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98324766.
RA OKANO M., XIE S., LI E.;
RT "Cloning and characterization of a family of novel mammalian DNA
(cytosine-5) methyltransferases."
RL NAT. GENET. 19:219-220(1998).
DR EMBL; AF068627; G3327982; -.
KW TRANSFERASE; METHYLTRANSFERASE.
SQ SEQUENCE 839 AA; 94798 MW; BD2ADD62 CRC32;

Query Match 54.5%; Score 54; DB 11; Length 839;
Best Local Similarity 46.7%; Pred. No. 7.76e-01;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 115 ESPAVTRHSGTSS 129
OY 1 QSGTMRTRHSGTGN 15

RESULT 8
ID 008509 PRELIMINARY; PRT; 859 AA.
AC 008509;
DB 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
GN DNA CYTOSINE-5 METHYLTRANSFERASE 3B1.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;

```

```

OC SCUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98324766.
RA OKANO M., XIE S., LI E.;
RT "Cloning and characterization of a family of novel mammalian DNA
(cytosine-5) methyltransferases."
RL NAT. GENET. 19:219-220(1998).
DR EMBL; AF068626; G3327980; -.
KW TRANSFERASE; METHYLTRANSFERASE.
SQ SEQUENCE 859 AA; 97232 MW; AECE323E CRC32;

Query Match 54.5%; Score 54; DB 11; Length 859;
Best Local Similarity 46.7%; Pred. No. 7.76e-01;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 115 ESPAVTRHSGTSS 129
OY 1 QSGTMRTRHSGTGN 15

RESULT 9
ID 080162 PRELIMINARY; PRT; 160 AA.
AC 080162;
DB 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
GN COMPLETE GENOME.
OS BACTERIOPHAGE BIL170.
OC VIRUSES; DSDNA VIRUSES. NO RNA STAGE; TAILED PHAGES; SIPHOVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RA CRUTZ-LE COQ A.M., CESSERLIN B., COMMISSAIRE J., ANBA J.,
RA KYRIAKIDIS S., CHOPIN M.C.;
RT "Sequence and organization of the lactococcal isometric bil170 phage
genome."
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF009630; G3282321; -.
SQ SEQUENCE 160 AA; 18072 MW; 67E4C50B CRC32;

Query Match 53.5%; Score 53; DB 9; Length 160;
Best Local Similarity 60.0%; Pred. No. 1.33e+00;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 90 GYMRNRHNG 99
OY 3 GTMRTRHSGT 12

RESULT 10
ID 038226 PRELIMINARY; PRT; 160 AA.
AC 038226;
DB 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
GN ORF3.
OS BACTERIOPHAGE BIL66.
OC VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95325326.
RA BIDENKO E., EHRLICH D., CHOPIN M.C.;
RT "Phage operon involved in sensitivity to the Lactococcus lactis
abortive infection mechanism AbiD1."
RL J. BACTERIOL. 177:3824-3829(1995).
DR EMBL; L35175; G845551; -.
SQ SEQUENCE 160 AA; 17986 MW; 9565A452 CRC32;

Query Match 53.5%; Score 53; DB 9; Length 160;
Best Local Similarity 60.0%; Pred. No. 1.33e+00;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINGOUGH R., ANDERSON K., BAYNES C., BEKES M.,
RA BONFIELD J., BIRTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAYTON M., DEAR S., DU Z., DUBBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,

RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPER A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.;
RL NATURE 368:32-38(1994).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA SAMMONS L., WOHLDMANN P., BIFMELD T.,
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.,
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL:AF038611; G2702418; -;
SQ SEQUENCE 844 AA; 95128 MW; E718BDA3 CRC32;

Query Match 53.5%; Score 53; DB 5; Length 844;
Best Local Similarity 58.3%; Pred. No. 1.33e+00;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
DB 737 ADAMRRKHSHTG 748
QY 2 SGTMRTRHSHTG 13

Search completed: Fri Jun 11 17:57:04 1999
Job time : 131 secs.